THE UNIVERSITY *of* EDINBURGH

MODEL-BASED ANALYSIS OF STABILITY IN NETWORKS OF NEURONS

DAGMARA PANAS

Doctor of Philosophy
School of Informatics
University of Edinburgh

2015

Dagmara Panas:

Model-based analysis of stability in networks of neurons

Doctor of Philosophy, 2015

SUPERVISORS:

Matthias H. Hennig

Luca Berdondini

LAY SUMMARY

Neurons, the building blocks of the brain, are an astonishingly capable type of cell. Collectively they can store, manipulate and retrieve biologically important information, allowing animals to learn and adapt to environmental changes. This universal adaptability is widely believed to be due to plasticity: the readiness of neurons to manipulate and adjust their intrinsic properties and strengths of connections to other cells. It is through such modifications that associations between neurons can be made, giving rise to memory representations; for example, linking a neuron responding to the smell of pancakes with neurons encoding sweet taste and general gustatory pleasure. However, this malleability inherent to neuronal cells poses a dilemma from the point of view of stability: how is the brain able to maintain stable operation while in the state of constant flux? First of all, won't there occur purely technical problems akin to short-circuiting or runaway activity? And second of all, if the neurons are so easily plastic and changeable, how can they provide a reliable description of the environment?

Of course, evidence abounds to testify to the robustness of brains, both from everyday experience and scientific experiments. How does this robustness come about? Firstly, many control feedback mechanisms are in place to ensure that neurons do not enter wild regimes of behaviour. These mechanisms are collectively known as homeostatic plasticity, since they ensure functional homeostasis through plastic changes. One well-known example is synaptic scaling, a type of plasticity ensuring that a single neuron does not get overexcited by its inputs: whenever learning occurs and connections between cells get strengthened, subsequently all the neurons' inputs get downscaled to maintain a stable level of net incoming signals.

And secondly, as hinted by other researchers and directly explored in this work, networks of neurons exhibit a property present in many complex systems called *sloppiness*. That is, they produce very similar behaviour under a wide range of parameters. This principle appears to operate on many scales and is highly useful (perhaps even unavoidable), as it permits for variation between individuals and for robustness to

mutations and developmental perturbations: since there are many combinations of parameters resulting in similar operational behaviour, a disturbance of a single, or even several, parameters does not need to lead to dysfunction. It is also that same property that permits networks of neurons to flexibly reorganize and learn without becoming unstable. As an illustrative example, consider encountering maple syrup for the first time and associating it with pancakes; thanks to sloppiness, this new link can be added without causing the network to fire excessively.

As has been found in previous experimental studies, consistent multi-neuron activity patterns arise across organisms, despite the interindividual differences in firing profiles of single cells and precise values of connection strengths. Such activity patterns, as has been furthermore shown, can be maintained despite pharmacological perturbation, as neurons compensate for the perturbed parameters by adjusting others; however, not all pharmacological perturbations can be thus amended. In the present work, it is for the first time directly demonstrated that groups of neurons are by rule *sloppy*; their collective parameter space is mapped to reveal which are the sensitive and insensitive parameter combinations; and it is shown that the majority of spontaneous fluctuations over time primarily affect the insensitive parameters.

In order to demonstrate the above, hippocampal neurons of the rat were grown in culture over multi-electrode arrays and recorded from for several days. Subsequently, statistical models were fit to the activity patterns of groups of neurons to obtain a mathematically tractable description of their collective behaviour at each time point. These models provide robust fits to the data and allow for a principled sensitivity analysis with the use of information-theoretic tools. This analysis has revealed that groups of neurons tend to be governed by a few leader units. Furthermore, it appears that it was the stability of these key neurons and their connections that ensured the stability of collective firing patterns across time. The remaining units, in turn, were free to undergo plastic changes without risking destabilizing the collective behaviour.

Together with what has been observed by other researchers, the findings of the present work suggest that the impressively adaptable yet robust functioning of the brain is made possible by the interplay of feedback control of few crucial properties of neurons and the general sloppy design of networks. It has, in fact, been hypothesised that any complex system subject to evolution is bound to rely on such design: in order to cope with natural selection under changing environmental circumstances, it

would be difficult for a system to rely on tightly controlled parameters. It might be, therefore, that all life is just, by nature, *sloppy*...

ABSTRACT

Neural circuits in all areas of the brain are subject to continuous remodelling due to the plasticity of individual synapses and neurons. Experience-driven learning, spontaneous fluctuations and homeostatic regulation all contribute to this ongoing process. However, despite the malleability of single cells, both behaviour and function of neural networks are generally found to be stable and reliable. This suggests that individual plasticity is somehow orchestrated across the network, in order not to destabilize the global functioning. To date, however, a mechanism for maintaining such group stability has not been identified. The present work was aimed to address this issue by using long-term *in vitro* recordings and subsequent model-based analysis of stability.

Repeated 15-minute recordings of cultured rat hippocampal neurons were obtained with high-density multi-electrode arrays over several days, revealing that *in vitro* preparations exhibit the desired global stability accompanied by local fluctuations. Subsequently, pairwise maximum entropy models were used to characterize activity patterns of groups of neurons within each separate recording. The particular analytic form of the employed model allowed for a quantification of parameter sensitivity in each group by using the Fisher Information Matrix. The analysis of obtained matrices has shown that the models exhibit two important properties: sloppiness and sparsity. Sloppiness entails a particular structure of parameter space, wherein the models are insensitive to changes in many of their parameters and parameter combinations. Sparsity, furthermore, results in the sensitive parameter combinations being aligned with only a few leader neurons. A comparison of group activities and model parameters over days further revealed that changes in those are related to sensitivity. In particular, it transpired that the vast majority of parametric changes occurred along the insensitive parameter subspace.

These results suggest that the few leader neurons, through being the most influential on group behaviour and at the same time the most stable over time, are responsible for global stability. At the same time, the remaining majority of less sensitive

neurons are free to explore the parameter space and undergo learning, without destabilizing the group behaviour.

ACKNOWLEDGEMENTS

Innumerable thanks to my supervisor Dr. Matthias Hennig, for all the ideas and inspiration, and for great patience and understanding. Most importantly, for giving me the freedom to explore and the help to find a way through, which is all a graduate student could wish for.

Many thanks to Dr. Luca Berdondini, my second supervisor, for the generous collaboration, hospitality in hosting me in his lab, and providing terabytes of valuable data, without which this work could not have been accomplished.

Many thanks also to Dr. Alessandro Maccione, who was an amazingly patient host and teacher in the lab and shared his expertise and time to help make this project happen (and shared also his expertise on Genoan cuisine, which should not be overlooked as a contribution to the creative process).

I am very grateful to Hayder Amin for all the time and effort he put in obtaining the data; he is solely responsible for the majority of successful recordings used in this work - not a trivial task, as I learned the hard way.

I am also very grateful to Oliver Muthmann, whose relentless work on improving spike detection was an important contributor to the quality of data I was to work with.

Many thanks go to Dr. Mark van Rossum, a source of excellent ideas, constructive criticism and illuminating brainstorming meetings.

A personal thank you to Greg for reading my thesis in full and giving me a very needed outside feedback, and helping me better transfer ideas onto paper.

I would also like to thank everyone at the Doctoral Training Center; it was a unique time and place to do a PhD, and I wouldn't have it any other way.

Last but not least, thanks to my folks, for infesting me with curiosity and for supplying me in honey throughout the years.

DECLARATION

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Edinburgh, 2015

Dagmara Panas

March 21, 2017

CONTENTS

1

INTRODUCTION

The brain is endowed with an impressive capability to learn and adapt to novel environments throughout its lifespan. This immense flexibility inherent to neurons allows not only for accumulation of information or precise body control but even for such feats of neuroengineering as direct neural control of prosthetic devices with the use of merely tens of cells (Ganguly and Carmena, 2009). Furthermore, what is perhaps even more impressive, these plastic changes do not interfere with already acquired skills, nor are they disrupted by further tasks learned by the very same neurons (Ganguly and Carmena, 2009). How is it possible that ensembles of cells can be collectively so functionally robust, when any given neuron retains the ability to undergo plastic changes at need?

This general question was the motivation for the work described in this thesis. In the present chapter, a broad introduction to the topic is provided, and the outline of the thesis is delineated.

1.1 MOTIVATION: PLASTICITY VERSUS STABILITY

The underpinnings of the adaptive capabilities of the brain lie in the fact that individual neurons are inherently plastic (Chklovskii et al., 2004). In fact, neuronal networks remodel continuously, not only as a result of experience-driven learning (Trachtenberg et al., 2002; Margolis et al., 2012), or in response to direct stimulation protocols (Fitzsimonds et al., 1997; Bi and Poo, 1999) or pharmacological perturbations (Nakayama et al., 2005; Slomowitz et al., 2015), but also driven by spontaneous ongoing activity (Maletic-Savatic et al., 1999; Okabe et al., 1999; Holtmaat et al., 2005; Minerbi et al., 2009; Margolis et al., 2012; Ziv et al., 2013).

Experiments employing *in vitro* cortical and hippocampal cultures demonstrate that most observable synapses change their size (Minerbi et al., 2009), and approximately 20% appear and disappear (Okabe et al., 1999) in the time course of several

days of unperturbed activity. Neurons in such preparations are also responsive to stimulation protocols that mimic learning, such as induction of long-term depression (Fitzsimonds et al., 1997) or potentiation (Bi and Poo, 1999) (interestingly, the changes in evoked potentials observed in those studies were not confined to the locus of stimulation but also affected distal synapses). Reports from *in vivo* studies also indicate that structural modifications that take place over periods of tens of hours and days result in a turnover of about 20% of synaptic boutons (Trachtenberg et al., 2002; Holtmaat et al., 2005). Such changes occur spontaneously (Trachtenberg et al., 2002; Holtmaat et al., 2005) and are further enhanced by sensory experience (Trachtenberg et al., 2002). Importantly, these appearing and disappearing boutons seem to be fully functional, since they form synaptic clefts and are equipped with synaptic vesicles (Trachtenberg et al., 2002). Moreover, a further receptive-field mapping of sensory-deprived neurons reveals that as their synaptic connectivity changes in result of the deprivation, so do their responses to stimuli (Trachtenberg et al., 2002). Indeed, particularly during learning, functional behaviour of single neurons is often modified from day to day (Lever et al., 2002; Carmena et al., 2005), with cells adjusting their tuning properties such as orientation preference (Carmena et al., 2005) or place field location (Lever et al., 2002). However, such adjustments do not necessarily require explicit training and also take place during ongoing activity, as cells change their sensitivity to stimuli (Margolis et al., 2012; Ziv et al., 2013).

Such ongoing plastic changes, if unchecked, would be expected to lead to unstable behaviour both in individual neurons and in neural ensembles (Turrigiano and Nelson, 2000), e.g. through disturbing the excitation-inhibition balance (Barral and Reyes, 2016), and to thus degrade stored information (Fusi and Abbott, 2007). Yet, when viewed collectively, networks of neurons exhibit remarkably consistent behaviour, maintaining the population firing rate distributions, complex activity patterns, and - crucially - their functional role in information processing and storage.

Cultured neurons, once they reach a mature state, exhibit stereotyped activity persisting over extended periods of time (Kamioka et al., 1996; Marom and Shahaf, 2002) and stable firing rates (Slomowitz et al., 2015). Furthermore, even though individual synapses change considerably, their overall distributions remain constant (Minerbi et al., 2009). Even more importantly, the functional representations encountered *in vivo* are found to be generally stable (Carmena et al., 2005; Chestek et al., 2007; Hu-

ber et al., 2012; Margolis et al., 2012; Ziv et al., 2013). For example, in animals trained on complex sensory-motor tasks, ensemble cortical activity becomes reliably predictive of behaviour (Huber et al., 2012). Similarly, monkeys performing motor tasks with the use of brain-machine interfaces develop stable cortical representations that can be consistently decoded from (Carmena et al., 2005). Also in the hippocampus the functional representation of environment relies on stable collective activity (Ziv et al., 2013). Interestingly, while in some studies global reliability of behaviour is attributed to stable function of most involved single units (Chestek et al., 2007; Ganguly and Carmena, 2009), others report opposing results (Carmena et al., 2005; Huber et al., 2012; Margolis et al., 2012; Ziv et al., 2013); that is, only some of the recorded neurons remain consistent. In the barrel cortex of mice, whisker tuning selectivity of cells has been observed to change in strength from session to session, in many neurons in fact exhibiting a reversal of selectivity: this occurs in 70% of low-responsive cells, and in 30% of high-responsive cells (Margolis et al., 2012). In another study, in mice trained to respond to whisker stimulation, single neurons in the motor cortex tended to largely retain preference to various aspects of the task (such as whisking versus licking); however, from session to session it was a different sub-population of neurons that contributed to a reliable ensemble representation (Huber et al., 2012). A similar effect was observed in the hippocampus, where place-coding ensembles of cells also exhibited a fluctuating membership, with only 15-20% of neurons overlapping between days (Ziv et al., 2013).

The above findings indicate that the brain is capable of maintaining stable functional representations despite considerable structural flexibility and ongoing fluctuations in the behaviour of individual cells. How is it that networks of neurons strike a balance between plasticity and stability?

It is now well established that experience-driven plasticity has a stabilizing counterpart in the form of homeostatic processes (a number of excellent reviews exist on the topic, focusing on various aspects of homeostatic plasticity, such as types and specificity, molecular and cellular mechanisms, modelling approaches etc: Turrigiano (1999); Turrigiano and Nelson (2000); Marder and Prinz (2002); Davis (2006); Marder and Goaillard (2006); Turrigiano (2011). This stabilizing form of plasticity is generally thought to operate according to the principle of negative feedback: whenever a designated biological signal indicates a departure from a homeostatic set-point,

e.g. a set average firing rate, the system counteracts this by adjusting its parameters to reverse the changes and maintain the desired target. For an illustrative example, both experimental and modelling evidence suggest that such a process takes place when the stomatogastric ganglion of the crab is deafferented, and thus deprived of excitatory neuromodulatory drive (Golowasch et al., 1999). Initially, the deprivation results in highly diminished activity of all cells and a cessation of the characteristic oscillatory pattern called the *pyloric rhythm*. However, within 48 hours the firing rates increase substantially without any further external intervention - in fact, sufficiently to re-establish the oscillatory behaviour of the intact system (Golowasch et al., 1999). In the model proposed to account for the recovery of the pyloric rhythm, internal calcium concentration plays the role of the negative feedback signal, since Ca reflects the activity level of neurons and is also known to be involved in signalling pathways related to gene transcription (Turrigiano, 2011), making it a likely candidate for a homeostatic sensor. The proposed mode of action in this scenario is re-establishing target cell firing rates through the regulating influence of calcium on ion channel conductances. This simple example, however, deals with the case where each neuron is driven back towards its original behaviour, and while it can explain certain phenomena, particularly in systems like the stomatogastric ganglion, where little learning plasticity is expected and high robustness is essential, it falls short when considering more complex cases.

In fact, a host of other signalling molecules have also been implicated to encode a target homeostatic set-point (such as brain-derived neurotrophic factor, Turrigiano and Nelson (2000)) and experimental observations indicate that in general a whole hierarchy of local homeostatic rules exist (Turrigiano, 2011), from the level of single synapses, through groups or types of synapses, to the level of whole-neuron synaptic scaling and changes to ion channel conductances. Furthermore, theoretical studies also propose non-local rules, whereby network activity is modulated in response to a putative diffusive signal such as nitric oxide (Sweeney et al., 2015). However, while this wide array of mechanisms provides the *means* for maintaining stability, what remains still unclear is how they are utilized in practice in various conditions and systems, and which aspects of neuronal structure and activity are under precise homeostatic control, while still ensuring that the desired individual plasticity is not counteracted (Turrigiano, 2011).

Since the evidence from homeostatic experiments (as discussed in earlier paragraphs) and the variability observed within and across individuals indicate that there can exist multiple network configurations resulting in the same operational regime (as discussed in (Davis, 2006; Marder and Goaillard, 2006; Blankenship and Feller, 2010)), this hints at the possibility of such mechanisms as different parameter changes compensating for one another, or redundancy of parameter sub-spaces. In such case, homeostatic plasticity would potentially be organized accordingly, e.g. ensuring another element of the system 'takes over' from the one that had undergone a change, or acting only on the non-redundant elements of the system.

It is instructive here to look to other fields faced with similar dilemmas. In the area of systems biology such dilemmas arise when considering that complex systems (be it gene regulatory networks or whole organisms) are robust to mutations and yet at the same time capable of evolving (Daniels et al., 2008). Indiscriminate negative feedback on every controllable parameter is not a feasible solution to the problem of robustness and conflicts with evidence, and so the interest in the field has grown in examining what is called *neutral spaces* (Daniels et al., 2008). Neutral spaces are the areas of parameter space that map onto the same behaviour of the system, e.g. different amino-acid sequences folding into a very similar and functionally equivalent enzyme, or gene combinations producing the same phenotype. The existence of such mapping provides a plausible explanation as to how perturbations of components of the system need not result in disturbances on a global scale and a break-down of functionality, and in fact might allow for evolving novel adaptive responses, such as a new binding active site of an enzyme. It transpires that in many cases in systems biology these neutral spaces are a result of a general property of many models, i.e. a highly anisotropic parameter space (Daniels et al., 2008), or *sloppiness*. That is, large regions of parameter combinations (termed *sloppy*) map onto very similar behaviour, however, there are a few directions in multiparameter space (termed *stiff*) along which the system is sensitive to changes. This property, together with properly applied negative feedback (i.e. ensuring that the few stiff parameter combinations are kept stable), provides a plausible mechanism for reconciling robustness and evolvability (Daniels et al., 2008; Transtrum et al., 2015).

In an analogous manner, such anisotropy in neural network models would explain how local individual plasticity need not result in global changes, and global stability

can be achieved through homeostatic regulation of the stiff subspace. The theory of sloppy systems mentioned above provides the tools to analyse multiparameter models and identify their sensitivity structure, thus allowing to directly test such hypotheses. Moreover, exploring parameter sensitivities can potentially be taken even further through sparsity analysis of discovered subspaces. Not to be confused with *coding sparseness*, an assertion in neural coding of sparse representation of stimuli by neurons and of low redundancy, here it is taken to mean that the stiff dimension itself is sparse, i.e. it is dominated by few parameter combinations. The distinction between a system that is merely sloppy and a system that in addition is sparse is a fundamental one - in the first case, collective stability can only be reached through simultaneous compensative homeostatic action on most parameters (e.g. a diffusive signal with opposing effects on complementary parameters), while the latter case implies that negative feedback can be selective due to redundancy (e.g. a subset of neurons with strongly conserved properties being sufficient to preserve global functionality). Distinguishing between those cases can have important implications in computational neuroscience, from understanding the hierarchy and interaction of homeostatic plasticity, through mapping the capacity of neural circuits to learn and adapt, through to answering questions on the nature of neural coding.

1.2 OUTLINE OF THE WORK AND THE ORGANISATION OF THE THESIS

The work described in this thesis was intended as basic research into the principles that underlie the relationship between group stability and the malleability of individual neurons. In order to conduct such investigations two things were required: an experimental system providing longitudinal data capturing the behaviour of many single neurons; and a model of neural activity amenable to parameter sensitivity analysis.

In Chapter 2, dissociated hippocampal neurons cultured on high-density microelectrode arrays are introduced as the experimental system. The choice of this particular neural preparation was dictated by several factors, as elaborated on in the Background of the chapter, along with a general overview of the various types of research conducted on those preparations. Briefly, *in vitro* cultures offer superior accessibility and reliability of single neuron monitoring as compared to most available

in vivo approaches; and the self-organized nature of dissociated preparations allows for abstracting away from particular properties of a given brain area, enabling more general conclusions about innate characteristics of neural networks and their emergent behaviour. In the Results section of the chapter, it is demonstrated that dissociated cultures meet the necessary criteria for the purpose of the study, as they exhibit globally stable behaviour, concurrent with ongoing changes in the activity of individual neurons. The chapter concludes with a Discussion of the findings in relation to other work and the limitations and strengths of employing dissociated cultures in the present context.

Chapter 3 introduces the pairwise maximum entropy model as the model of choice for dissecting the interplay of local and global behaviour of neurons. In the Background section, recent research in this area is summarized, illustrating the insights that can be gained from this type of modelling. Importantly, the pairwise maximum entropy approach provides interpretable parameters, an analytically tractable mathematical form, and has been demonstrated to capture well the correlated behaviour of groups of neurons. In the Results section, it is shown that also in the current setting these models provide good fits to the experimental data; it is also demonstrated that even in such simplified description of neural statistics it is not straightforward to understand the interplay of parameter fluctuations and global stability. Finally, in the last section the similarities and differences to other reported research in the area are discussed.

In Chapter 4, the framework of sloppy theory is employed to test the main hypothesis of the thesis through the analysis of parameter sensitivity of the model. A broad outline of research on sloppiness is provided in the Background section, highlighting the universality of this approach. In the Results section the main findings of this work are reported, congruent with the posed hypothesis. It transpires that indeed the modelled neurons exhibit the property of sloppiness, and that the fluctuations observed in the activity of individual neurons tend to predominantly occur along the insensitive parameters. Furthermore, and importantly, the stiff dimensions of modelled cells are shown to be sparse and align roughly with individual units, suggesting an interpretation where *hub* neurons provide a stable backbone for group activity. In the Discussion section the results are compared to those of other studies and the robustness of the findings is addressed.

Chapter 5 presents an attempt to test the translatability of the above results to an *in vivo* setting. In the Background section the limitations of *in vitro* investigations are delineated, as well as challenges of examining stability in the living brain. In the Results section, analysis of a publicly available *in vivo* recording is presented, employing the framework of maximum entropy modelling and sloppy theory-based analysis. The findings are not aligning with those from dissociated cultures; however, this is confounded by the fact that the chosen recording is possibly not suited for the particular methodology employed here. In the Discussion section these negative findings are put in perspective of the particular challenges posed by the chosen data-set and of the general challenges of *in vivo* experiments.

In the concluding chapter of the thesis a general discussion of presented research is provided: relation to other work and broader implications, the various limitations and possible improvements, as well as inspiration for further investigations in this domain.

2

MONITORING NEURAL ACTIVITY ON HIGH-RESOLUTION MULTIELECTRODE ARRAYS

2.1 BACKGROUND

Dissociated neuronal cultures are a common and well-established tool in neurophysiological research, with over thirty years' history of use and a wide range of applications (broadly reviewed by Verderio et al. (1999) and Marom and Shahaf (2002)). Research performed in these preparations spans many spatial and temporal scales, as well as different levels of abstraction, from investigations of basic neural and synaptic physiology (Verderio et al., 1999), through various aspects of neuronal plasticity (Fitzsimonds et al., 1997; Bi and Poo, 1999; Okabe et al., 1999; Minerbi et al., 2009; Arnold et al., 2005), development (Kamioka et al., 1996; Han and Stevens, 2009), circuit physiology (Bettencourt et al., 2007), pharmacology (Kamioka et al., 1996; Arnold et al., 2005; Han and Stevens, 2009; Minerbi et al., 2009), through to tentative attempts at understanding the nature of neural code (Shahaf et al., 2008) and sleep function (Hinard et al., 2012). Among various *in vitro* model systems of brain-like activity (extensively reviewed in Corner (2008)), cultures offer a simple yet powerful approach (Marom and Shahaf, 2002): they are relatively easy to maintain in a stable condition for long periods of time; they offer superior accessibility for both recording, stimulation, pharmacological interventions and optical imaging; and they reflect primarily the innate properties of neurons, since they are allowed to self-organize without external input.

Dissociated cultures are obtained through plating tens of thousands of pre- or neonatal neurons on the surface of microelectrode arrays (MEAs, see Figure 2.1) coated with binding proteins (Marom and Shahaf, 2002). The cells are supplied with necessary growth factors and glucose in the form of liquid medium, which allows them to develop and interconnect (Marom and Shahaf, 2002). Over the first few days *in vitro* (DIV), the neurons mature, extend axons and dendrites, and begin synapse

Figure 2.1: A recording chip such as used for this study, combining multielectrode array (light-gold square in the middle) and a culturing dish (glass ring around the electrode area). Photograph courtesy of the manufacturer, 3Brain GmbH.

formation (Kamioka et al., 1996; Verderio et al., 1999; Marom and Shahaf, 2002; Wagenaar et al., 2006). Importantly, some of the neurons form coupling with the electrodes, which allows for subsequent recordings of their electrical activity (Kamioka et al., 1996; Segev et al., 2001, 2004; Chiappalone et al., 2006; Wagenaar et al., 2006; Shahaf et al., 2008; Berdondini et al., 2009b). Within the first week *in vitro*, action potentials can be recorded, at first asynchronous and sparse (Kamioka et al., 1996; Wagenaar et al., 2006), over next week becoming more widespread and progressively more synchronized (Kamioka et al., 1996; Wagenaar et al., 2006), and finally reaching a stage of culture-wide bursts of activity interspersed with extended periods of relative quiescence (Kamioka et al., 1996; Segev et al., 2001; Marom and Shahaf, 2002; Segev et al., 2004; Wagenaar et al., 2006; Chiappalone et al., 2006). This developmental time course reflects the fact that the neurite growth phase lasts until around second week *in vitro*, when connections are spanning the whole culture (Wagenaar et al., 2006); and synapse formation outnumbers synapse elimination until the beginning of the fourth week (Okabe et al., 1999; Nakayama et al., 2005). At this point, cultures are considered to reach maturity (Nakayama et al., 2005), characterized by a relatively stable pattern of bursting (Segev et al., 2001; Marom and Shahaf, 2002; Chiappalone et al., 2006; Segev et al., 2004) and an established robust homeostatic response to manipulation (Kamioka et al., 1996; Han and Stevens, 2009). At the same time, neurons retain

their natural plasticity, as documented by both structural (Okabe et al., 1999; Minerbi et al., 2009) and functional studies (Marom and Shahaf, 2002; Slomowitz et al., 2015). Such mature cultures can then be recorded from for months (Kamioka et al., 1996; Marom and Shahaf, 2002), allowing to observe tens (Kamioka et al., 1996; Wagenaar et al., 2006; Chiappalone et al., 2006), hundreds, and even thousands (Berdondini et al., 2005, 2009a,b) of individual neurons simultaneously. Importantly, the morphological signatures and various aspects of activity resemble in many ways the activity in the brain, which is what makes dissociated cultures a highly useful tool (Marom and Shahaf, 2002).

Bursting in particular has received a lot of attention, since this complex behaviour has counterparts in *in vivo* observations (Corner, 2008). In one of the early studies of MEA-grown cortical cultures, Kamioka et al. (1996) recorded neural activity over several weeks *in vitro* and used pharmacological agents to establish whether spontaneous activity plays a role in development. They first identified three distinct stages of electrical activity: asynchronous spiking (3-10 DIV); early, regular synchronous bursts (11-17 DIV); and mature, irregular synchronous bursts (31-38 DIV). The authors then reversibly silenced the spiking activity with tetrodoxin (TTX, a potent toxin that prevents the generation of action potentials by binding to the voltage-gated sodium channels of neurons) at a range of *in vitro* stages, to test whether lack of activity is detrimental to maturation. Indeed, the induced lack of spontaneously generated action potentials during early, but not mature, stage, resulted in bursting patterns differing from the control. Authors concluded that 'spontaneous regular synchronized bursting is essential for normal development of synaptic connections and of neuronal networks' (Kamioka et al., 1996).

Extensive literature has since developed on the topic, with a breadth of applications and a variety of analysis methods. A particularly interesting course has been set by Segev et al. (2004), who analysed correlations between bursts and used clustering to test whether there is any structure in obtained correlation matrices. They discovered that bursts can be distinguished as belonging to separable classes. Authors further established that the neuronal activation patterns underlying each class of events are distinct, thus expanding on earlier studies which had established a lack of single point of origin for bursts (as reviewed in Marom and Shahaf (2002)). Furthermore,

the neuronal activities within different bursts were suggestive of underlying rank ordering.

A study by Shahaf et al. (2008) explicitly tackled that idea, by investigating whether order-based representation can be used for stimulus classification. The authors used electrical stimulation pulses at 2-6 different MEA sites as separate stimuli, and measured the population response (including only reliably responding neurons). They adopted the order of activated neurons as neural code and computed the distances between all obtained responses from two stimuli. They then tested supervised and unsupervised learning algorithms based on the above distance matrix, in both cases successfully classifying the neural responses with respect to the stimulus. They further extended these results by comparing the order-based representation with a spike-timing based representation in a supervised learning classification task, showing that for a realistic range of working parameters they perform similarly.

Demonstrably, dissociated neuronal cultures provide a testing ground for a range of scientific questions, and a system suited for experimental manipulations with access to direct responses of whole neuronal populations. It is therefore important to devote some resources to understand cultures as a system in its own standing, as in the work of Wagenaar et al. (2006).

In one of the most extensive studies to date Wagenaar et al. (2006) followed 58 cortical cultures of varying plating densities over the span of 4-5 weeks *in vitro*, augmenting the MEA recordings with electrical stimulation. In accordance with literature, authors observed the typical developmental course, from early neurite extension and sparse activity, through increasing synchrony, leading up to mature, complex bursting patterns. However, authors uncovered that the plating density profoundly affected the development, with denser cultures reaching bursting faster than the sparser ones. Based on the electrical stimulation experiments at different culture ages, the authors suggested a link between the development of synchronous bursts and the development of long-range connections. The authors also emphasised the rich diversity of bursting behaviours observed, varying not only with plating density, but also between sister and non-sister platings (i.e. from the same or a different animal), and between different culture dishes from the same animal preparation. Intriguingly, this diversity also extended to single cultures, with some specimens changing their bursting pattern even in the 5^{th} week *in vitro*. At this age all the main developmental events

– like neurite extension and major synaptogenesis – are complete, therefore such observations raise questions about the extent of stability and plasticity in mature neural networks.

Several studies directly address this issue by various approaches. For example, Minerbi et al. (2009) developed a system where dissociated neurons were plated on transparent MEA dishes so that the recording chips could be put under a microscope. Thus the authors were able to observe synapse turnover in individual neurons via fluorescent tagging of a well-known protein of the post-synaptic density (a protein-dense region specialized to the postsynaptic membrane), and at the same time record the spiking activity of the network. Results indicated that, after an initial increase in both activity, number of synapses and breadth of synapse size distributions, the cultures stabilized and maintained stable overall distributions of synapses. However, this was not accompanied by individual stability of identified fluorescent puncta, as these kept changing all throughout experiments. Authors further elucidated that the ongoing synaptic remodelling was partly activity driven, but also exhibited an intrinsic 'drift'. Altogether, an important conclusion is that even at their mature state, cultures are still highly plastic, with ongoing spontaneous changes in structure at the level of individual neurons.

Further insights into the topic come from a recent study by Slomowitz et al. (2015), who probed whether the behaviour of single neurons remains stable along with population stability. They examined cultured hippocampal neurons by means of MEA activity monitoring, as well as intracellular recordings, carried out over several days at around 3^{rd} week *in vitro*. Confirming the general assertion of stable mature state of the network, the authors found that the firing rate distributions across the culture, as well as global synchrony measures, remained stable throughout experiments. In contrast, single units did not maintain their individual firing rates. When drug intervention was applied to probe the mechanisms of homeostatic plasticity, the firing rate distributions and global synchrony measures returned to control levels on the second day. However, similarly to control conditions, single neurons behaviour was not maintained, with firing rates significantly changing. Furthermore, the excitation to inhibition ratio also changed, as did the frequency and amplitude of miniature excitatory post-synaptic potentials. Altogether, this has led the authors to challenge the cell-autonomous homeostatic rules, and demonstrated that some – as yet unknown

– global homeostatic mechanisms must be in operation, even in a model system as simple as dissociated cultures.

It appears therefore that dissociated cultures comprise a particularly useful model system for investigating the interplay between collective stability and local malleability of neurons, which is the purpose of the present study. First, these preparations meet the necessary requirements of exhibiting stable global behaviour concurrent with substantial individual plasticity; and the ongoing changes reflect the innate, basic properties of neurons, not confounded by a complex interaction of processes encountered *in vivo*. Second, they allow for monitoring tens and hundreds of neurons with single-unit precision and excellent time resolution, providing a comprehensive picture of network activity over days of experiments; something that is currently difficult to achieve *in vivo* (Lütcke et al., 2013). This enables a detailed model-based investigation of the relationship between global stability and local fluctuations.

2.2 MATERIALS AND METHODS

The neural activity recordings used in this work were performed by Hayder Amin, Alessandro Maccione, and the author in the laboratory of Luca Berdondini, with the materials provided by the Italian Institute of Technology in Genova. All procedures involving experimental animals were approved by the institutional IIT Ethic Committee and by the Italian Ministry of Health and Animal Care (Authorization ID 227, Prot. 4127 March 25, 2008). The recording protocol was chosen by the author under consultation with Alessandro Maccione, Luca Berdondini and Matthias Hennig. The algorithms used for spike detection were developed and implemented by Oliver Muthmann (paper in preparation), in C++ and Python.

All data processing and statistical analyses, as well as the parametric and methodological choices pertaining to those, were the responsibility of the author. Unless stated otherwise, the data analyses and figure preparation were performed with the use of custom scripts written in Matlab.

2.2.1 Dissociated neuronal cultures: preparation and maintenance

The neurons for dissociated cultures were obtained from brain tissue of prenatal Sprague-Dawley rats (both sexes) at embryonic day 18. To that end, hippocampi of 3 rat embryos were finely chopped, exposed to a trypsin solution, then mechanically dissociated and re-suspended in Neurobasal medium. Subsequently, 35,000-40,000 cells were plated on microelectrode array chips pre-treated with adhesion factors, poly-L-lysine and polyethyleneimine. Following established lab protocols, the cultures were incubated in 1ml Neurobasal medium supplemented with 1% Glutamax and 2% B-27 (all media substrates manufactured by Invitrogen), in a humidified atmosphere with CO2 to O2 ratio of 5 : 95, at 37°C. To ensure healthy development and maintenance of activity, 50% of the medium was changed every 3-4 days.

Cultures were recorded from the 24^{th} day *in vitro* (DIV) onwards, in order to ensure that all developmental processes were complete and preparations had established stable activity. In general for dissociated cultures, such maturity is typically considered to occur by the 3^{rd}-4^{th} week *in vitro* (Kamioka et al., 1996; Chiappalone et al., 2006; Wagenaar et al., 2006). Prior to recordings, all cultures were visually inspected under the microscope to confirm the presence of healthy cell bodies and to check for signs of pathology (Figure 2.2).

As the recordings took place outside of the incubator, for the duration the on-chip dishes were closed with sterilized custom-designed polydimethylsiloxane (PDMS) caps. The caps prevented media evaporation and osmolarity changes while permitting gas exchange. They additionally served to keep the neuronal preparations sterile. After each recording, the PDSM caps were removed under a sterilized hood and the chips were then placed in the incubator.

2.2.2 High-resolution multielectrode array recordings

High-resolution extracellular recordings were performed on the BioCam4096 platform with the use of the Active Pixel Sensor (APS) microelectrode array (MEA) chips (Berdondini et al., 2005), type BioChip 4096S (3Brain GmbH, Switzerland). These devices provide 4096 microelectrodes of 21μm$\times 21\mu$m in size, arranged in a 64×64 grid with inter-electrode separation of 21μm, on an active area of 2.67mm$\times$2.67mm

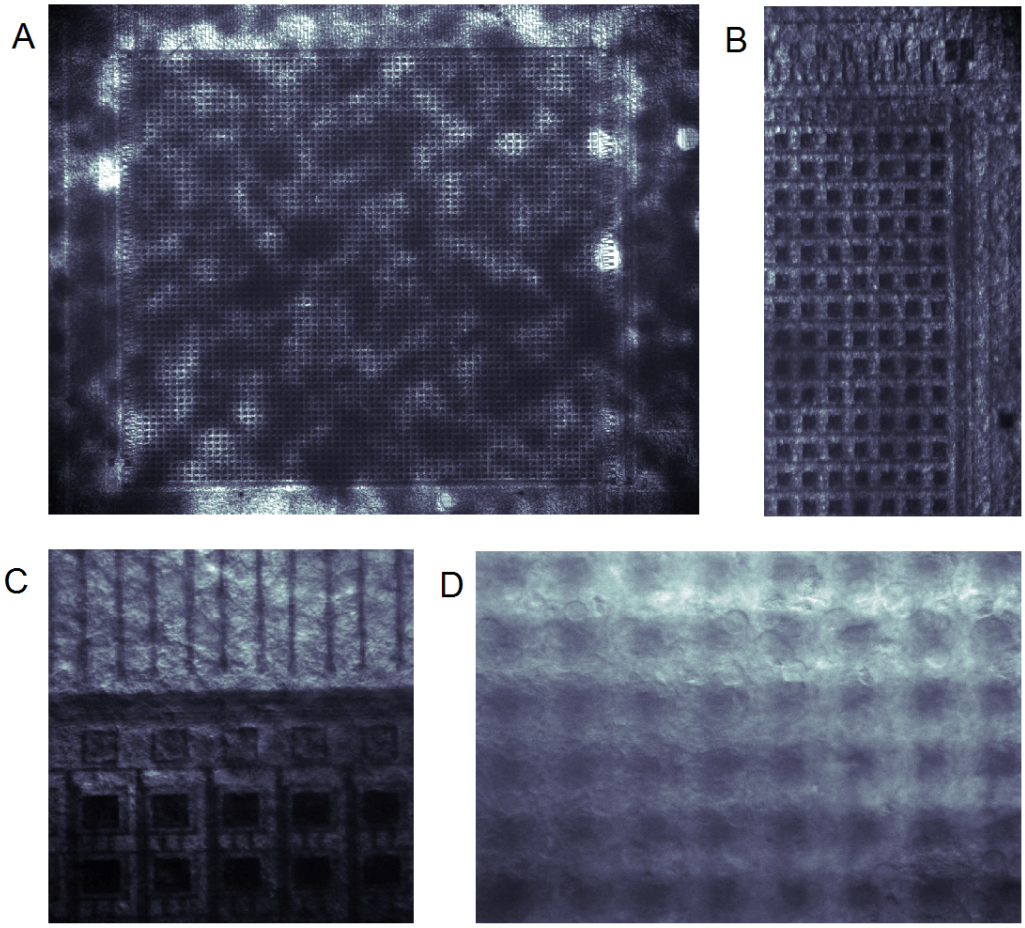

Figure 2.2: (A) Differential interference contrast microscopy image of a hippocampal culture plated on the APS microelectrode array; 5x magnification with the whole 64×64 array in field of view. Varying shading indicates uneven distribution of tissue, with darker areas corresponding to greater thickness of cell cover. Neural tissue exhibits signs of healthy growth, densely covering the chip and extending beyond the active area. (B) The same culture under 20x magnification. Electrodes are 21μm x 21μm in size, spaced 21μm apart. (C) Border region under 50x magnification in high focus. At this magnification neuronal bodies can be visible; in high focus they are more readily observed in the upper part of the image, where the background is lighter. (D) Central region under 50x magnification in low focus. Over the active area, neurons further from the chip surface are easier to observe. The size of cell bodies in the plane of focus is comparable to the size of the electrodes.

(Figure 2.2). Such high spatial resolution can be achieved by the combination of complementary metal-oxide-semiconductor (CMOS) technology with the on-chip in-pixel voltage amplification and, to deal with the data load, with the addressing logic adopted from image processing (Berdondini et al., 2005, 2009a). However, such amount of data puts limits on temporal resolution: when measuring from the full 64×64 channel array, the platform records at a sampling rate of 7.022 kHz per electrode. Here the full array was employed, in order to observe as many neurons as possible, building on earlier studies that established a reliable electrode-to-neuron correspondence in this recording system (Berdondini et al., 2009b).

Activity was recorded at 12 bits resolution per channel, low-pass filtered at 5 kHz with the on-chip filter and high-pass filtered by setting the digital high-pass filter of the platform at 0.1 Hz. Raw data were visualized and stored using the BrainWave software application provided with the BioCam4096 platform.

Recording duration was set to 15 minutes. This was dictated by two factors: health of the cultures; and practical limitations of data storage and recording scheduling. Even though the platform does not overheat the chip during recordings and the PDSM caps prevent media evaporation, it was decided to err on the side of caution, and limit the time and frequency of handling the cultures outside of the incubator. Second, the raw file of a 15 minute recording takes up approximately 50 GB, and data needs to be transferred periodically to external disks.

Recordings were performed twice on the first day of experiments and then daily, as reported in Table 2.1. These intervals between recordings were, to an extent, exploratory, although the choice was motivated by the observations of natural fluctuations in synaptic contact turnover over the span of hours, as reported by Okabe et al. (1999) and Minerbi et al. (2009).

	0h	2h	20h	44h	50h	68h	92h
culture 1	x	x	x		x	x	x
culture 2	x	x	x	x			
culture 3	x	x	x	x			

Table 2.1: Recording points for each of the cultures.

2.2.3 Spike detection: algorithms

Spikes were detected with a custom threshold-based algorithm, augmented with correlation-based analysis, both implemented by Oliver Muthmann. Full details of these methods are available upon request, as the research is currently pending peer review. Here a short summary is provided, with the consent of the author.

First, putative events were identified as negative voltage deflections crossing a specified threshold and satisfying simple shape constraints. The threshold t was defined in units of noise estimate n relative to the baseline b; that is, voltage deflections below baseline greater than t times the noise level. Baseline and noise estimates were updated in each time frame as follows:

- b was increased by $1/4\ n$ if the raw signal exceeded $b + n$ and decreased by $1/2\ n$ if the raw signal was below $b - n$; this effectively provided a running measure of the local voltage tertile;
- n was decreased by $1/64$ current voltage value if the signal was within the interval $[b, b - n)$ or (to compensate for spikes) exceeding $b - 6n$, and increased by the same amount for signal amplitudes $[b - n, b - 5n)$.

These baseline and noise estimates were used, instead of the commonly used mean and standard deviation, because the distributions of raw voltage values were broad and non-Gaussian. Therefore, percentile-based estimates were deemed a more robust choice to detect sparsely occurring spikes of voltage.

Second, following threshold detection, putative events were accepted as potential spikes if their shape satisfied the following criteria:

- peak width of at least 2 frames;
- no other minimum in the subsequent 7 frames (approximately 1ms);
- at least 2 of the subsequent 7 frames rising above baseline.

These criteria reflect the generally accepted features of spike morphology, i.e. a minimal duration, existence of a repolarization phase and existence of a refractory period. Third, in an offline post-processing stage, spike amplitudes were re-thresholded

to remove global fluctuations over the chip, and coincident spikes from neighbouring electrodes were removed to leave only the highest-amplitude one.

Since it is impossible to obtain the ground truth data on spiking for many neurons simultaneously, spike detection algorithms for MEAs are notoriously difficult to optimise with respect to the parameters used. As a result, in practice arbitrary choices need to be made with regard to, e.g. the detection threshold. This results in a trade-off between the amount of falsely rejected and falsely accepted events. In the spike detection used here, in an attempt to mitigate this issue, a novel approach was proposed: coupling low-threshold detection with further statistical analysis of correlations and amplitudes (referred to later on as correlation analysis). The low threshold decreased the incidence of false negatives, and the subsequent correlation-based analysis allowed for a data-driven identification of false positives.

The post-detection analysis exploited the fact that spikes are typically at least weakly correlated due to network interactions, whereas electrode noise is random and uncorrelated. First, to identify correlated channels across the array, electrodes with elevated activity were chosen as reference units. Using only such a subset of channels for comparison, rather than all available electrodes, served to increase efficiency and reduce the influence of noise from sparsely active channels. Then, spike times of all reference channels with an added clocking signal were ranked. These ranks were then used to determine the size of the time window to be used for the assessment of correlation; here it was 200 ranks. This variable time window was used in order to remove the confounding effect of fluctuating firing rates. Then, to determine whether a channel was correlated with a reference unit or not, the number of co-occurring spikes between those channels within the correlation window around each spike was compared to what would be expected from random coincidence. For random sequences, the expected number of events in an interval before and after a spike is given by a Poisson distribution; therefore, a channel was classified as correlated with the reference unit if the number of co-occurring spikes yielded a p-value < 0.1 under the Poisson probability distribution.

Second, for each putative spike on each electrode all co-active channels from the correlated pool were identified. This provided the basis for a definition of a correlation index (CI) of each putative spike: the fraction of events in the correlation window coming from correlated channels, multiplied by the fraction of significant correlations

among those channels. Thus the CI quantified the correlation of each event with the overall activity. This index was also computed for generated random events, which were inserted into each channel (Poisson noise with a rate of 0.25 Hz). Then, the cumulative distribution of CIs in each channel was compared against the corresponding cumulative distribution of surrogate random events. Based on this comparison, true positive events were identified as those with statistically significantly different CIs from the noise. Once the estimated fraction of false positives was established, a probability-like measure p_c was then assigned to each individual event, factoring in both the CI value of the putative spike and the overall reliability of the channel.

The performance of the algorithms was verified by analysis of an empty chip recording, where raw data was obtained from a MEA chip filled with physiological fluid but not plated with neural tissue. Since electrode noise results in false detections, even on an empty chip, the data consisted solely of false positives. Correlation analysis was shown to correctly reject most of the detected events as noise, without the need to increase amplitude threshold (which in a real dataset would result in false negatives).

2.2.4 Spike detection: choice of detection parameters

The set of algorithms described above did not entirely eliminate the trade-off between false negatives and false positives, nor the need for a choice of detection parameters. However, using a suitably low amplitude threshold and factoring in the information from correlation analysis allowed for this choice to be more data-driven and thus improved detection, as illustrated below.

The threshold used in the first step of detection was set to $t = 4.5$. As shown in Figure 2.3, this resulted in a skewed distribution of amplitudes, rapidly increasing from around threshold value. Importantly, the threshold was low enough so that there was no cut-off at low amplitudes, which ensured minimising the number of false negatives at this stage.

Next, after re-thresholding the amplitudes in reference to global fluctuations and removing coincident events, correlation analysis was applied. This procedure yielded a per-event probability-like measure, quantifying the likelihood of it being a true positive. Figure 2.4 shows the distributions of amplitudes and probability measures p_c of

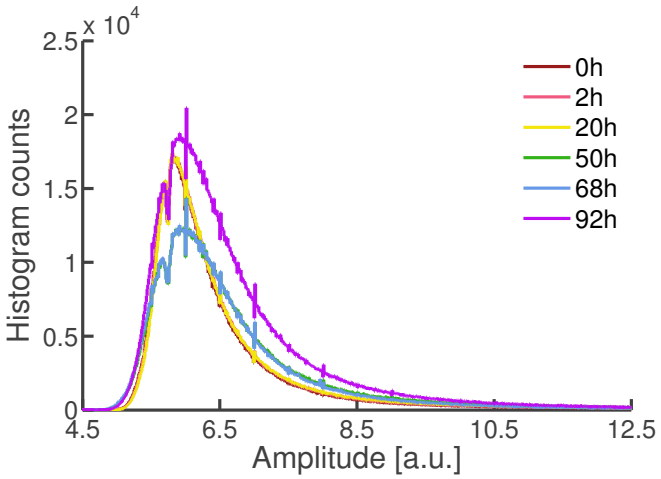

Figure 2.3: The distributions of amplitudes of detected events before the application of post-processing in culture 1 for each of the recordings.

detected spikes from each recording in culture 1. The distributions of amplitudes are clearly unimodal, without any indication of a boundary between noise and spikes. In contrast, the distributions of probability values are bimodal, with a sharp peak at low probabilities, and a broad skewed peak of intermediate values with a long high-probability tail. This form of event probability distribution suggests an underlying mixture of two types of events with partially overlapping distributions: uncorrelated noise, and true spikes. Analysis of an empty chip recording, mentioned in Section 2.2.3, corroborates this observation and the false positives stemming from electrode noise are classified as low-probability events.

Therefore, upon visual inspection of the distributions, a threshold of $p_t = 0.1$ was chosen to accept events as true spikes for further analysis. To further illustrate the benefits of this approach to spike detection, a comparison of the effects of using a moderate amplitude threshold ($t = 6$) versus the probability measure p_c threshold ($p_t = 0.1$) is presented below.

Firstly, Figure 2.5 depicts the number of detected events at each electrode of the 64×64 lattice, segregated into sub- and suprathreshold for each of the threshold criteria used. As can be seen in panel A, using the amplitude threshold results in an uneven distribution of false positives over the array, with clusters mirroring the layout of the presumed true activity (see panel B). This would suggest either the presence of signal-correlated noise, or that amplitude threshold erroneously rejects true low-amplitude activity. In contrast, the layout of false positives under the p_c threshold

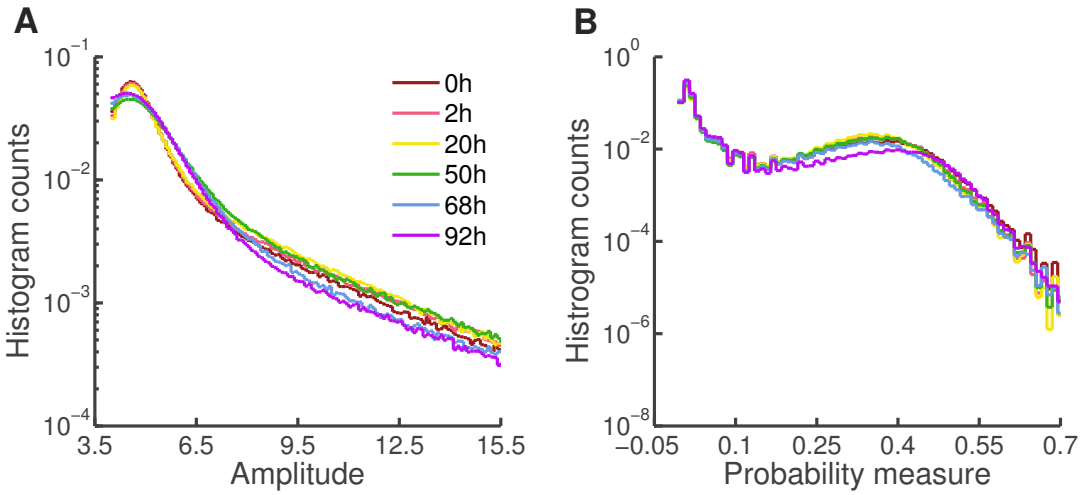

Figure 2.4: The distributions of amplitudes and the distributions of probability measures of detected events in culture 1 for each of the recordings. (A) Normalized histogram counts of the amplitudes of detected events for each recording, after amplitudes were re-thresholded and coincident spikes removed. (B) Normalized histogram counts of the probability measures of detected events for each recording. Log ordinate scale used for better clarity.

is more uniform (see panel C), as would be expected from random electrode noise. Indeed, the uniform distribution of such noise was confirmed with the analysis of an empty chip recording, where false positive detections resulted in a homogeneous distribution over the recording area, and a few isolated noisy electrodes.

Secondly, Figure 2.6 shows the histogram counts of the detected events across all electrodes in 5*ms* time bins, again segregated into sub- and suprathreshold categories. As evident from comparing panels A and B, the distribution of false positives under the amplitude threshold is inhomogeneous in time, with highest counts of false positives coinciding with the culture-wide bursts of activity. Again, this would suggest either signal-correlated noise or erroneous classification of spikes as false positives. Inspecting the recorded voltage traces indicates that the latter is a more likely explanation. Neurons in culture exhibit characteristic population bursts that consist of a mixture of single spikes from non-bursting neurons and volleys of spikes from bursting neurons (Wagenaar et al., 2006). This results in a wide variety of amplitudes and spike shapes being recorded at the times of bursts (Wagenaar et al., 2006). In particular, this also results in variability of amplitudes even from single neurons, which might be contributing to increased false positive rate at the times of bursts.

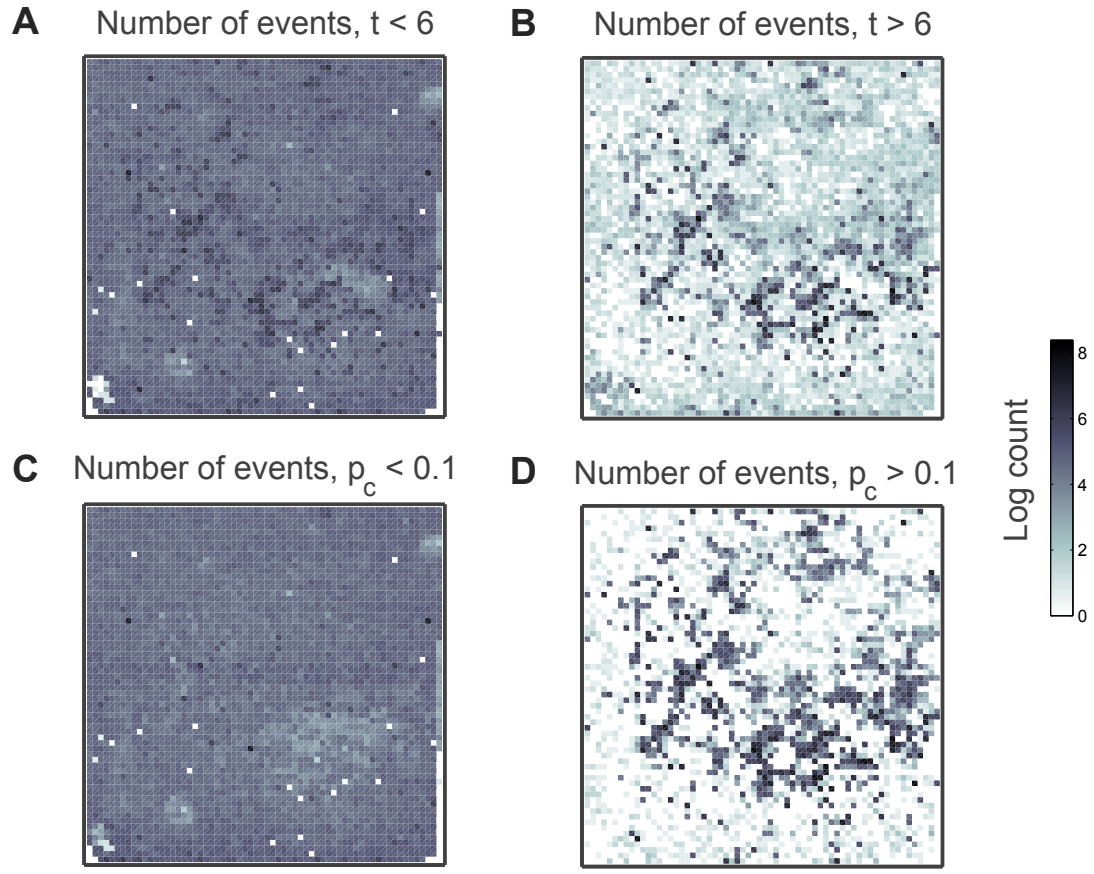

Figure 2.5: Comparison of the distributions of sub- and suprathreshold detected events on the multielectrode array for the two types of threshold considered. (A) Number of detected events at each electrode with an amplitude < 6, arranged on a 64×64 lattice representing the MEA layout. Data from culture 1, baseline recording. (B) Number of events with amplitude > 6 for the same recording. (C-D) As in A and B, number of sub- and suprathreshold detected events, here using a probability measure threshold of 0.1.

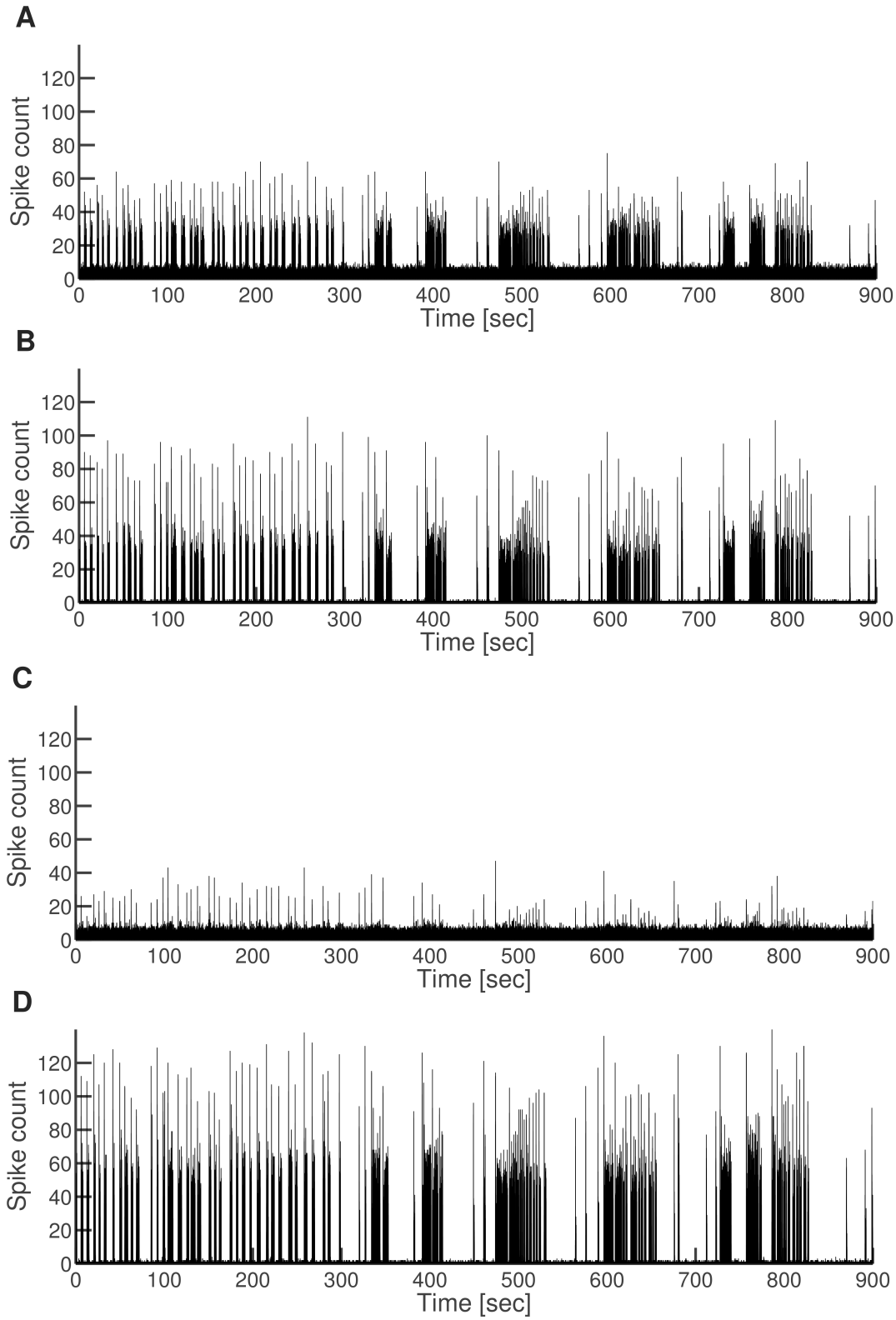

Figure 2.6: Comparison of the spike count over time of sub- and suprathreshold detected events for the two types of threshold considered. (A) Histogram of detected events with an amplitude < 6, in $5ms$ time bins. Data from culture 1, baseline recording. (B) Histogram of events with amplitude > 6 for the same recording. (C-D) As in A and B, number of sub- and suprathreshold detected events, here using a probability measure threshold of 0.1.

Taken together with the fact that amplitude-only detection exhibited clustered sub-threshold activity co-localized with putative neurons, this suggests that amplitude criterion resulted in considerable amount of false negative detections. In contrast, employing low amplitude threshold and correlation analysis allowed a substantial fraction of such events to be detected.

Finally, only channels with firing rates higher than 0.1 Hz were considered for further analysis, to ensure adequate number of spikes for statistical tests and modelling (here and throughout, firing rates refer to average firing rates over the whole recording).

2.2.5 Verification of single unit monitoring

Since the aim of this work is to examine the relationship between individual and global stability, it is important to verify consistent monitoring of single units across recordings. As has been established in earlier studies carried out with high-density MEAs by (Berdondini et al., 2009b), the weak signal spread between electrodes and small electrode size result in a one-to-one correspondence between recording channels and single neurons. Only in rare cases more than one neuron is recorded from one contact, since the electrode size is comparable to the size of neuronal somata (and in those rare cases such spikes are removed, see Section 2.2.4). Also rare are the cases of cross-talk between electrodes, due to the properties of shunt resistance in densely connected neural tissue of *in vitro* preparations (Berdondini et al., 2009b). However, the stability of such one-to-one correspondence has not yet been studied on the APS system. Other studies (Slomowitz et al., 2015) report that motility of mature hippocampal neurons is low, and mechanical shifts are not likely to affect neuron placement once the tissue is interconnected and anchored to the electrodes with the coating agents. To further verify reliable single neuron monitoring in the present study, spike shape similarity between recordings was examined.

The spike detection algorithm described in Section 2.2.3, in addition to writing out the time stamps of presumed spikes, also provided short cut-outs (21 frames, i.e. approximately 3ms) of the voltage signal. Once spikes were detected and channels with insufficient activity were filtered out, the mean spike shape within a recording was obtained for each electrode. These average spike shapes could then be compared be-

tween recordings, as illustrated in Figure 2.7, using the Pearson correlation coefficient as a similarity measure.

Spike shapes were consistently very highly correlated between consecutive recordings, with mean and median across all channels and in all preparations always above 0.98. Only a few outliers had a coefficient of correlation below 0.9, such as the first pair of shapes in the example presented in the bottom panel of Figure 2.7, where it appears that the recorded spike has changed between sessions. However, in the overwhelming majority of cases the spike shapes remained stable according to the correlation coefficient.

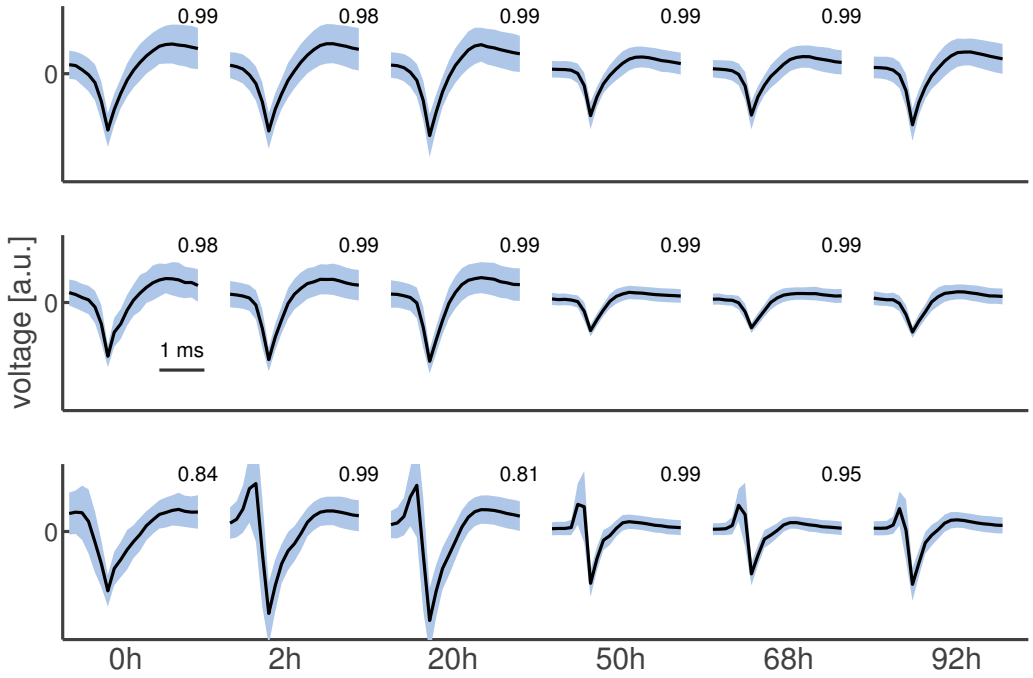

Figure 2.7: Example comparisons of the mean spike shape across recordings for three electrodes; data from culture 1. Coloured region corresponds to standard deviation. Top panel shows the electrode with highest coefficient of correlation between mean spike shapes of the first and second recording. Middle panel depicts a randomly chosen electrode. Bottom panel shows the electrode with the lowest coefficient of correlation between mean spike shapes of the first and second recording. The scaling down of voltage traces observed between 20 and 50h is most likely an artefact of the media wash carried out between those two recording points, as discussed later in the main text.

Examining the illustrated examples, it is in some cases apparent that indeed the spike shape retained its particular characteristics, such as e.g. a triphasic profile or an after-spike depolarizing inflection. However, comparing the first and second panels of Figure 2.7 it can also be noticed that in general the shape of a spike as recorded

from extracellular medium with relatively low temporal resolution is a fairly stereotyped physiological signal. Therefore, to obtain a frame of reference for understanding whether the similarity between recordings is particular to individual neurons or rather a generic property of spikes, correlations of spike shapes between different electrodes within a recording were also computed.

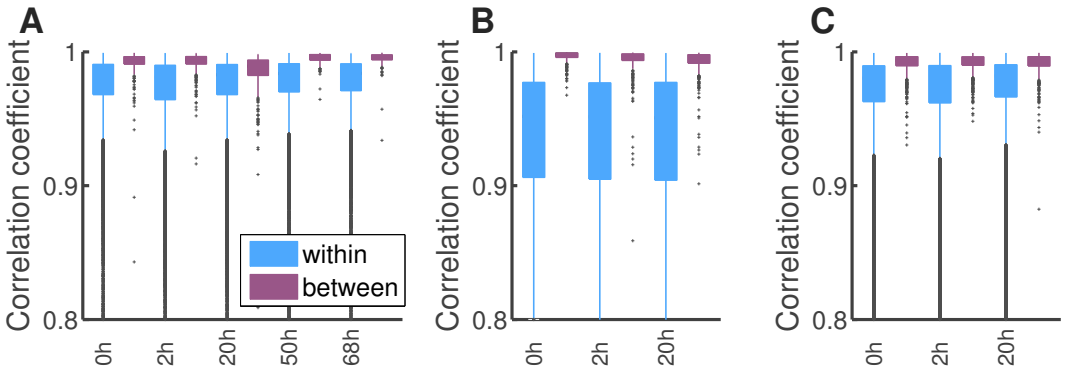

Figure 2.8: Comparison of the distributions of spike shape correlation coefficients within and between recordings. (A) Boxplots of the Pearson correlation coefficients for culture 1. (B-C) Same for cultures 2 and 3.

As evident from comparing the distributions of correlation coefficients within a single recording versus between consecutive recordings (see Figure 2.8), even though spike shapes can appear very similar at distinct electrodes, they are significantly more similar at a single electrode at different time points. This is especially prominent in recordings from culture 2, shown in Figure 2.8B, where the variety of spike waveforms was greatest of all recorded cultures. In contrast, it can be noticed that in culture 1 the correlation coefficients between shapes recorded at 20h and 50h are significantly lower than in all other cases. This, however, was likely an effect of the media change carried out at 45h, and not a reflection of a decrease in the level of correspondence of single neurons between recording sessions. As can be seen in Figure 2.7, all channels were similarly affected and reported lower amplitudes after as compared to before the media change, while still retaining other characteristics of its own spike shape.

In conclusion, the stability of single unit monitoring appeared to be preserved, and from here on the terms neuron, unit and channel are used interchangeably.

2.2.6 Estimating correlations

For both theoretical and practical considerations, the focus of the study is unit spiking activity, and therefore correlations were calculated based on binned spike trains. To guide the choice of appropriate time bin, two approaches were considered: estimating the half-width of the cross-correlograms of voltage traces between example units; and literature research of relevant time scale of synaptic events. The first approach suffers from several issues, most notably varying levels of recording noise between preparations and the necessity to rely on an arbitrary subset of available channels (since it is computationally infeasible to compute cross-correlograms for many units). In light of this, the time bin choice was informed by literature research. As reported by Fitzsimonds et al. (1997) and Erickson et al. (2008), the majority of monosynaptic latencies between a pre- and postsynaptic spike in a dissociated hippocampal culture fall within 1-5 ms, while polysynaptic connections had an average latency of 9ms (Erickson et al., 2008). Since the aim of the study is to better understand the relationship between individual traits of neurons and group behaviour, and including polysynaptic connections in the analysis would have confounded the distinction between two-neuron and multi-neuron interactions, only monosynaptic connections were of interest. Therefore, a time bin of 5ms was chosen for all further analyses.

Spike trains were binned into binary vectors, with 1 corresponding to spike and -1 corresponding to silence. Correlations were then calculated as the sample Pearson correlation coefficients between each pair of active channels. To assess the significance of correlations, two-sided hypergeometric test was used.

2.3 RESULTS

In total, 14 high-resolution recordings of extracellular potential were obtained from 3 healthy hippocampal non-sister cultures. The recorded voltage traces were analysed to detect action potentials as described in Section 2.2.3, and then channels with firing rates over 0.1 Hz were taken to represent active neurons. To provide an idea of the nature of activity under study, sample raster plots of detected spikes are presented in Figure 2.9, and a close-up of a 40-second time frame from culture 1 is presented in Figure 2.10.

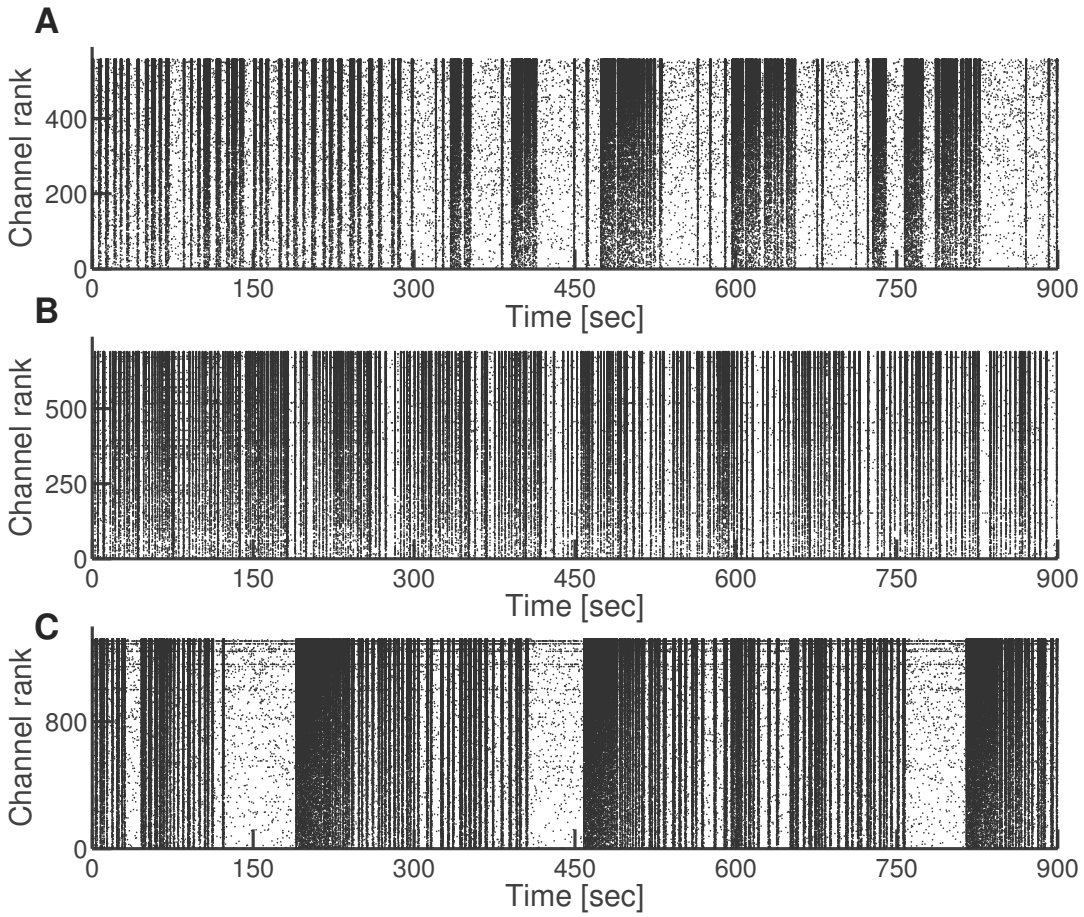

Figure 2.9: Illustration of the typical activity of dissociated neurons, with prominent bursts of synchronized spiking. (A) Raster plot of individual spikes in each channel with channels ranked by their firing rate; each dot denotes a spike. Baseline recording at 0h in culture 1. (B-C) Same for cultures 2 and 3, also from baseline recording.

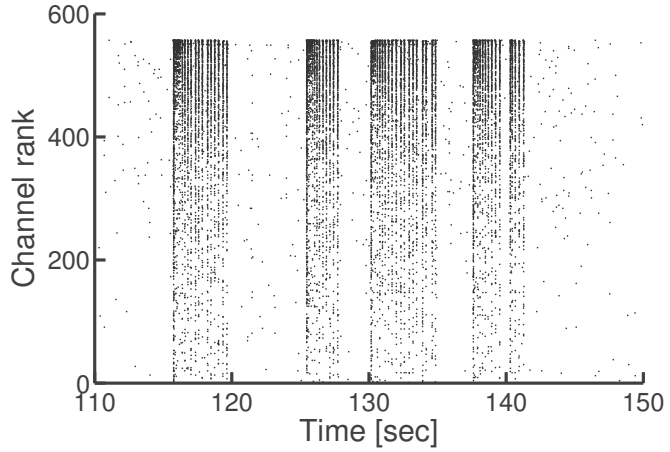

Figure 2.10: A close-up of a few of the bursts in culture 1, highlighting the sparse nature of the spiking activity.

This activity is typical of dissociated cultured preparations, with very sparse activity (all mean and median firing rates below 1Hz) and highly synchronized bursts of activity that recruit the majority of firing units into brief wave-like events (Kamioka et al., 1996; Segev et al., 2001; Wagenaar et al., 2006; Corner, 2008).

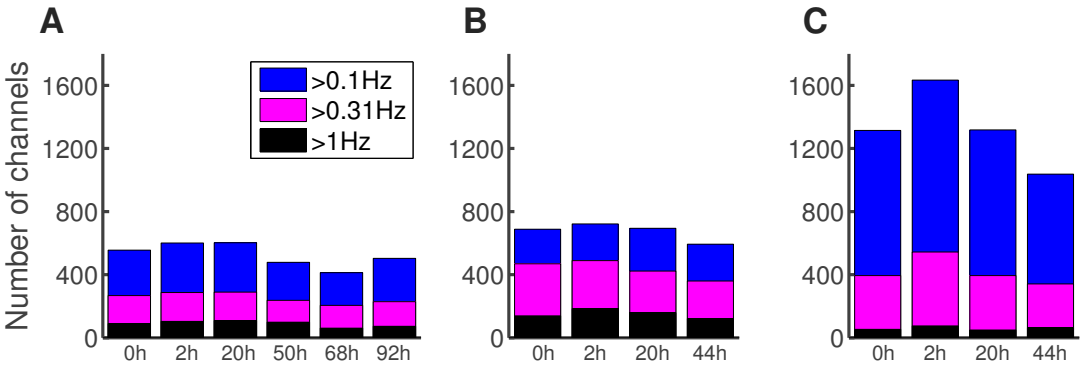

Figure 2.11: Numbers of recorded active channels (over 0.1Hz) across preparations and recordings. (A) Bars representing the number of active channels in each of the recordings from culture 1. (B-C) Same for cultures 2 and 3.

The numbers of reliable active channels differed substantially between preparations, as shown in Figure 2.11. Such variability is expected between networks of dissociated neurons (Wagenaar et al., 2006; Corner, 2008), even when each culture is prepared in the same manner and the same cell density is used for plating; it is a result of the interplay between random wiring and the genetic and epigenetic properties of cells, in absence of the normalizing developmental stimuli from the rest of the brain (Corner, 2008).

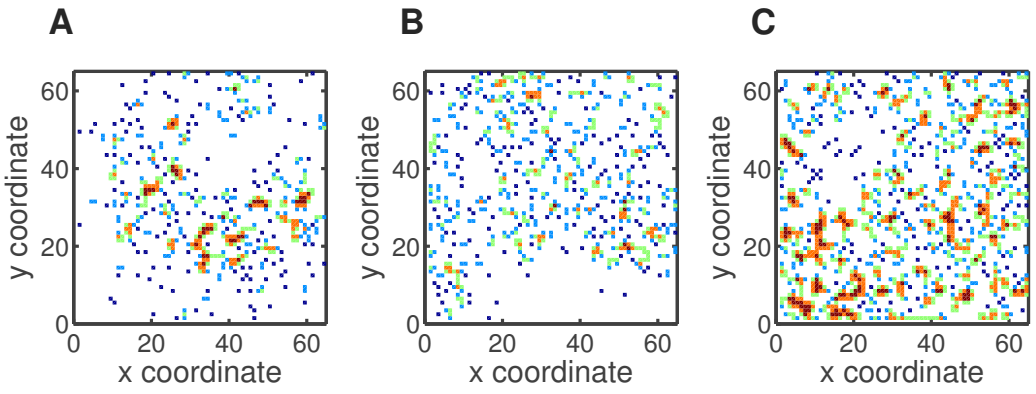

Figure 2.12: Varying structure of the layout of reliable active channels over the array from different recordings. (A) Active channels in culture 1, baseline recording at 0h; color code denotes how many nearest neighbours are also active, to provide a visual guide to how clustered the activity is. (B-C) Same for cultures 2 and 3, also from baseline recording.

As mentioned earlier, the variations in random wiring contribute to the inter-preparation variability; and indeed this is reflected in the differing layout of cells over the array, as illustrated in Figure 2.12. In the case of cultures 1 and 3, there was a clear tendency of active electrodes to form clusters (as illustrated by the warm colors in Figure 2.12). In culture 1, 22% of all active electrodes had at least 3 active neighbours, and in culture 3 this percentage was even higher, with 35% of active electrodes being clustered in this manner. On the other hand, culture 2 exhibited noticeably more sparse and homogeneous spread of active electrodes, with most active channels having one or two neighbours, and only 10% having 3 or more.

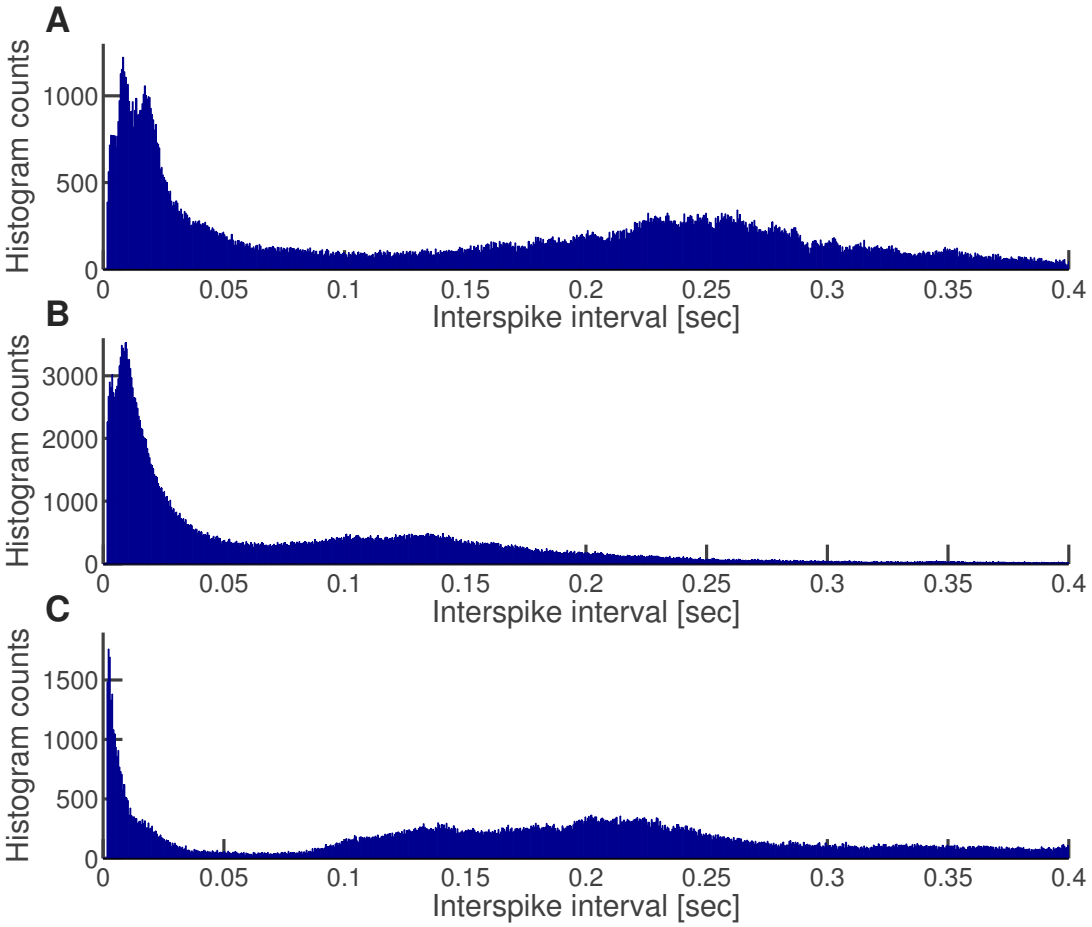

Figure 2.13: Example distributions of interspike intervals. (A) Histogram counts of pooled interspike intervals of all channels in culture 1, baseline recording at 0h. The size of the time bin for histogram was 0.5ms. (B-C) Same for cultures 2 and 3.

The inter-preparation variability is also well documented in other measures, most notably a broad range of burst properties (Wagenaar et al., 2006) that dissociated

cultures of neurons can exhibit. Visual inspection of example raster plots of spiking activity in Figure 2.9 indicates that the cultures used for this study also appear to exhibit a noticeable degree of heterogeneity in the manner of their bursting. This is also confirmed by examining the interspike interval distributions from each culture (see Figure 2.13).

Despite the inevitable inter-preparation variability, the activity of dissociated cultures is in many ways stereotyped and so was the data used here. First, in all cultures and recordings the distributions of firing rates were highly skewed towards low values. Only a fraction of units exhibited spiking frequencies above 1 Hz (Figure 2.11, black; see also Figure 2.14A, C and E in the next subsection) and the mean firing rates were always lower than 0.72Hz (Table 2.2). Even though the least active channels were filtered out, skewness coefficients were always higher than 1.5 (Table 2.3).

	0h	2h	20h	44h	50h	68h	92h
culture 1	0.53	0.57	0.63		0.71	0.57	0.55
culture 2	0.64	0.74	0.66	.62			
culture 3	0.32	0.34	0.31	0.36			

Table 2.2: Mean firing rate values [Hz] in each of the recordings.

	0h	2h	20h	44h	50h	68h	92h
culture 1	3.23	3.11	3.25		4.53	4.33	3.72
culture 2	1.526	1.61	1.59	1.57			
culture 3	5.1	6.1	7.2	10.1			

Table 2.3: Sample skewness (i.e. third central moment divided by the cube of standard deviation) of the firing rate distributions in each of the recordings.

Second, despite apparent global synchrony of bursts, the pairwise channel correlations between binned spike trains were very low (see next section Figure 2.14B, D and F), with mean Pearson correlation coefficient always lower than 0.06 (Table 2.4). However, these correlations were predominantly significant – as assessed by a test against a hypergeometric distribution at $\alpha = 0.05$, the fraction of insignificant correlations was lower than 0.1 in most recordings (Table 2.5).

To conclude, in general terms the collected data appears to constitute a representative sample of spontaneous cultured activity, without indications of atypical outliers, but at the same time exhibiting a range of behaviours expected from random differences between preparations.

	0h	2h	20h	44h	50h	68h	92h
culture 1	0.046	0.041	0.044		0.047	0.06	0.05
culture 2	0.047	0.043	0.051	0.057			
culture 3	0.033	0.025	0.033	0.03			

Table 2.4: Mean values of the pairwise Pearson correlation coefficient between binned spike trains in each of the recordings.

	0h	2h	20h	44h	50h	68h	92h
culture 1	0.055	0.054	0.047		0.073	0.034	0.043
culture 2	0.066	0.088	0.078	0.063			
culture 3	0.16	0.105	0.205	0.097			

Table 2.5: Fractions of all calculated correlations in each recording that did not reach statistical significance.

2.3.1 Population stability

As stated earlier (Section 2.2.1), the time point of 24^{th}DIV was chosen to ensure that all developmental processes are complete, since the aim of this study was to investigate stable global activity. In order to confirm that the cultures exhibited the desired stability, the properties of population-level behaviour were compared between recordings and across cultures. For visual inspection, the cumulative distributions of firing rates and the histograms of Pearson correlation coefficients between binned spike trains are presented in Figure 2.14.

As can be observed from comparing panels A, C and E, the firing rate distributions exhibit a degree of preparation specific features which are preserved across recordings. This does not appear to be the case with correlations (panels B, D and F of same figure). In order to further assess this, statistical analyses were performed.

First, the Lilliefors test of normality was administered to the distributions of firing rates and the distributions of correlations, to guide the choice of appropriate statistical measures and methods for similarity assessments. In all cases, the null hypothesis of normality was rejected with $p < 0.001$ and therefore nonparametric tests were employed throughout. Also, since the distributions were skewed and non-normal, median population values were compared, rather than the means.

The median values of firing rate distributions are presented in Figure 2.15. To estimate the confidence intervals (CI) of the median in each recording, a bootstrap

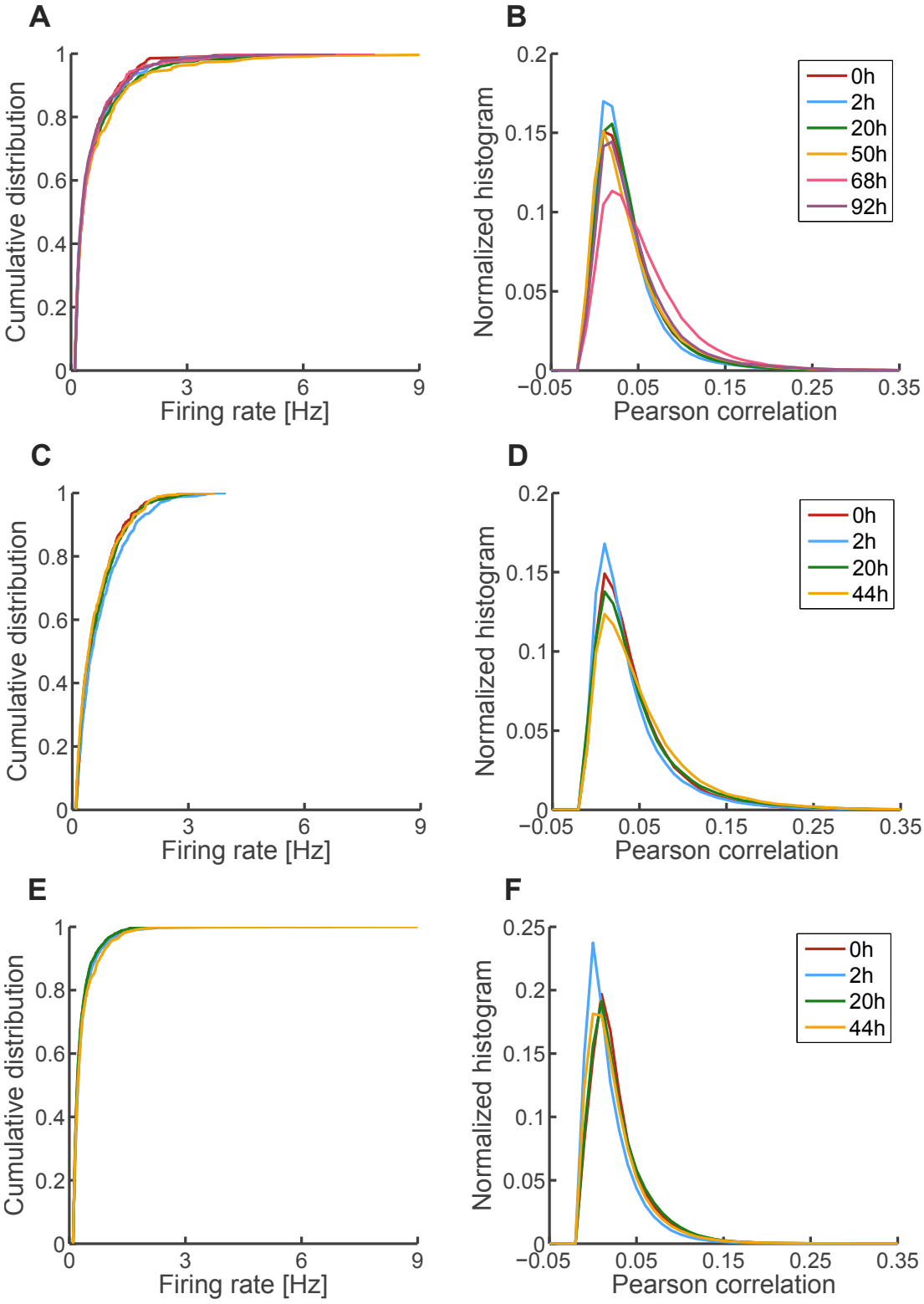

Figure 2.14: Distributions of firing rates and correlations across recordings. (A) Cumulative distribution functions of firing rates in culture 1 from all recordings; color code shown in panel B. (B) Histogram bin counts of Pearson correlation coefficients between all pairs of channels in culture 1 (Pearson correlation coefficients calculated for spikes binned into 5ms time bins). Bin size used for histogram: 0.01. (C-D) (E-F) Same for cultures 2 and 3.

approach was employed, with 100000 iterations of re-sampling the firing rates. The 2.5 and 97.5 percentiles of the obtained distribution of median values were used as CI. Results indicate that the medians of the distributions did not change significantly from recording to recording, with the exception of culture 3. The confidence intervals in this preparation were smaller than in the others, likely reflecting the higher number of active channels. However, this was the culture with the highest skewness coefficients (see Table 2.3), which also may have contributed to overestimated CI.

What is also worth noting is that the median firing rates were always significantly different between the preparations (note different ordinate ranges in Figure 2.15), indicating that inter-culture variability was pronouncedly higher than the within-culture variability.

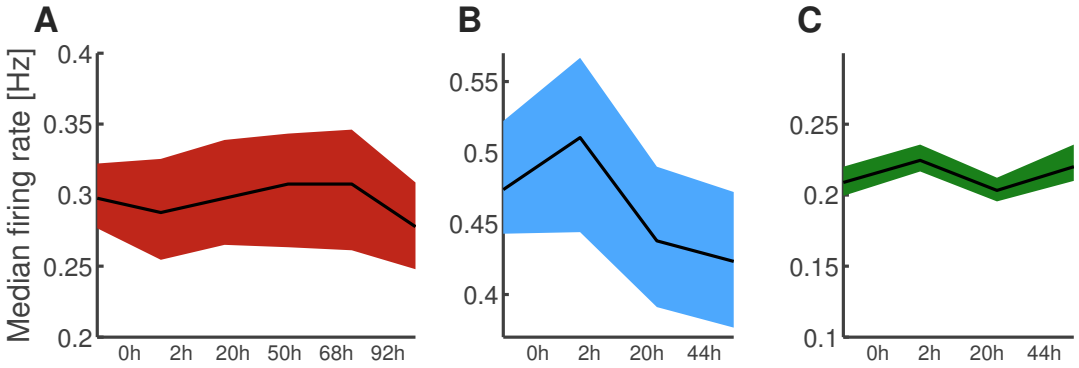

Figure 2.15: Median firing rates across recordings. (A) Median firing rate plotted as a function of recording, with confidence intervals estimated by bootstrap; culture 1. (B-C) Same for cultures 2 and 3.

To further examine the stability of distributions, also the skewness coefficients were compared between consecutive recordings (Figure 2.16), using the same bootstrap approach as above. Confirming earlier general observations, all distributions of firing rates were significantly skewed (that is, skewness coefficients were significantly different from null). In cultures 1 and 2 the skewness coefficients were stable within-culture and significantly different between-culture. However, in culture 3 the estimated CI were so large as to preclude statistical differentiation from other cultures or between recordings. Together with the results of median CI estimation, this suggests either a lower degree of firing rate distribution stability in culture 3, as compared to cultures 1 and 2.

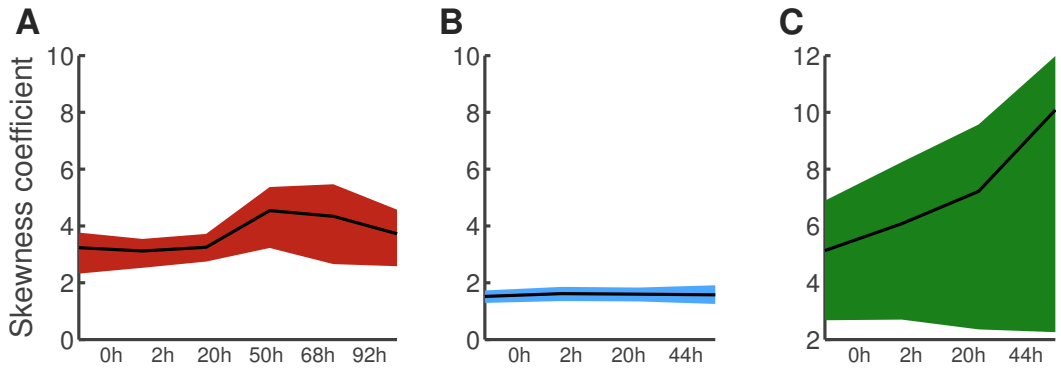

Figure 2.16: Skewness coefficients of the firing rate distributions across recordings. (A) Skewness coefficient plotted as a function of recording, with confidence intervals estimated by bootstrap; culture 1. (B-C) Same for cultures 2 and 3.

As a complementary approach to comparing distribution medians and skewness coefficients, to test the null hypothesis of whether firing rate distributions from consecutive recordings were drawn from the same underlying population, Kolmogorov-Smirnov goodness-of-fit test was administered to each pair of recordings. The resulting p-values for the firing rate distributions, reported in Table 2.6, indicate that in most cases the null hypothesis cannot be rejected at $\alpha = 0.01$. In fact, if Bonferroni correction for multiple testing was to be used, that would hold for all cases and $\alpha = 0.05$ (apart from the 2-20 comparison in culture 3). However, it needs to be noted that a failure to reach significance in a statistical test does not constitute a proof of the null hypothesis. Therefore, to put those results in context, goodness-of-fit test was also administered to distributions of firing rates between the baseline recordings from the different cultures. The p-values for these tests were all several orders of magnitude lower than the lowest reported p-values for within-culture comparison ($p < 10^{-10}$).

	0-2h	2-20h	20-44h	20-50h	50-68h	68-92h
culture 1	0.303	0.466		0.857	0.176	0.701
culture 2	0.064	0.021	0.56			
culture 3	0.022	0.004	0.071			

Table 2.6: The p-values of the Kolmogorov-Smirnov test between distributions of firing rates from consecutive recordings.

The p-values of the Kolmogorov-Smirnov test for the distributions of correlations, on the other hand, were significant, with p-values lower than 10^{-5} in all comparisons. Interestingly, these values were comparable to – and sometimes lower than – values

obtained from inter-culture comparisons. The median correlation coefficients were also significantly different between consecutive recordings (not shown). Together with visual assessment of distributions of correlations, this would suggest that: all cultures were likely in a stable regime of behaviour; however, the distributions of correlations cannot be treated as an individualized marker of behaviour, preserved across days in a statistically stringent manner.

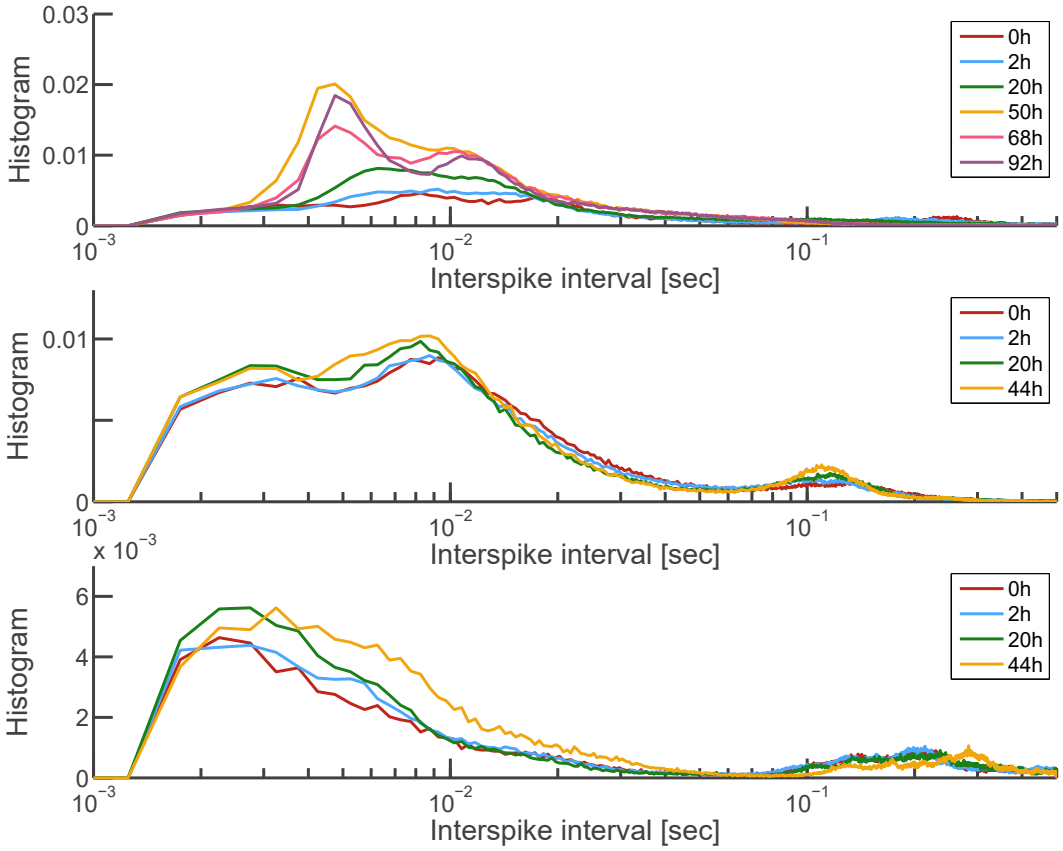

Figure 2.17: The distributions of interspike intervals across recordings. (A) Normalized histogram counts of the pooled interspike intervals of all active channels in culture 1; different recordings color-coded as in legend. The size of the time bin for histogram was 0.5ms. (B-C) Same for cultures 2 and 3.

Finally, to explore a different point of view on stability of overall behaviour, the histograms of interspike intervals were compared between consecutive recordings and across the baseline recordings from different cultures. As presented in Figure 2.17, the interspike interval distributions appeared to remain fairly stable from recording to recording in majority of the cases. When Pearson correlation coefficients were calculated between the histogram count distributions, these were in most cases higher than 0.95 (see Table 2.7). In contrast, when compared between recordings, the correlation

coefficients were always lower than 0.9 (cultures 1-2: 0.857, 1-3: 0.551, 2-3: 0.633). Furthermore, these results hold across a range of histogram bins tested (0.1ms, 0.25ms, 0.5ms, 1ms, 5ms). It should be noted, however, that in culture 1 there appears to be a shift in the distribution shape between the 20h and the 50h mark. As mentioned earlier in Section 2.2.5, between those two recordings a media wash was administered. It appears that this intervention, in addition to affecting global voltage noise levels (see Figure 2.7), also may have affected the bursting patterns in the preparation. However, the culture remained stable before and after the wash.

	0-2h	2-20h	20-44h	20-50h	50-68h	68-92h
culture 1	0.938	0.946		0.92	0.983	0.986
culture 2	0.995	0.992	0.995			
culture 3	0.976	0.956	0.892			

Table 2.7: Pearson correlation coefficients between the distributions of interspike intervals in consecutive recordings.

Overall, it appears that collected data was widely stable, with distributions of firing rates and interspike intervals largely preserved between consecutive recordings, to a degree significantly above the inter-preparation variability.

2.3.2 Changes in behaviour of individual neurons

While the overall behaviour of cultures can be considered to be in a stable, mature regime, this does not preclude significant fluctuations in the activity levels of single neurons. As an illustrative example, in Figure 2.18 the changes in individual firing rates and the changes in average correlations of single neurons are plotted, arranged on a square grid corresponding to the layout of the electrodes. In fact, given the observations from other studies (Okabe et al., 1999; Minerbi et al., 2009; Slomowitz et al., 2015), and the general plasticity of neurons, it was expected that at least a degree of such changes would be observed on an individual level. To examine whether this was the case, here individual parameters of single neurons were compared, rather than distributions. To that end, common active channels were identified between consecutive recordings.

First, as a counterpart to the Kolmogorov-Smirnov statistical tests reported in Section 2.3.1, a nonparametric test for paired samples was performed, the Wilcoxon

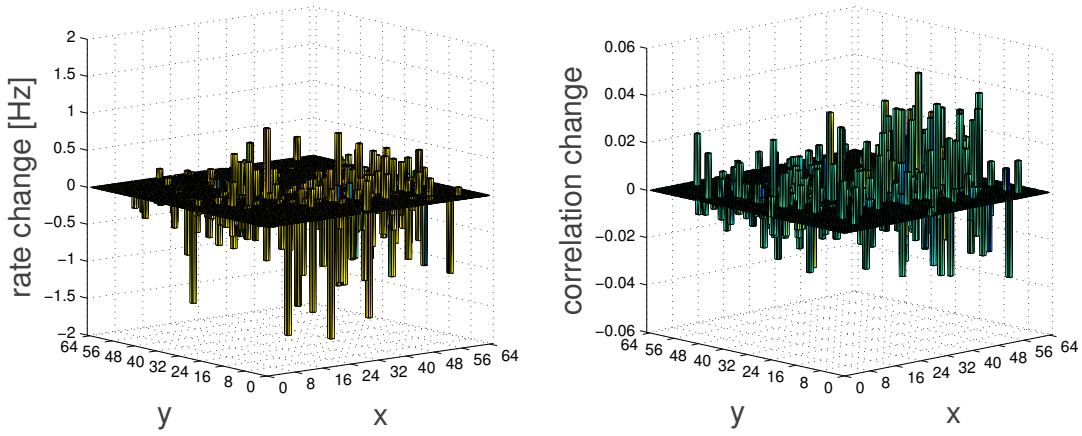

Figure 2.18: Illustration of activity changes between recordings in individual neurons. (A) The change in firing rate of each individual unit between recordings at 0h and 2h, culture 1, arranged on a grid layout. (B) The change in the average Pearson correlation coefficient of each individual unit with all other active channels; culture 1, between 0h and 2h.

signed rank test. Thus, it could be tentatively examined whether the population stability was simply due to the stability of individual firing rates. The resulting p-values are reported in Table 2.8. These results indicate that, from a statistical point of view, the individual firing rates could not be considered pairwise-stable, with the exception of a borderline case of last pair of recordings in culture 3.

	0-2h	2-20h	20-44h	20-50h	50-68h	68-92h
culture 1	< 0.0001	< 0.0001		0.001	< 0.0001	< 0.0001
culture 2	< 0.0001	< 0.0001	< 0.0001			
culture 3	< 0.0001	< 0.0001	0.016			

Table 2.8: The p-values of the Wilcoxon signed rank test between paired populations of firing rates from consecutive recordings.

However, the Wilcoxon test relies on ranks, and thus ignores the magnitudes of the changes in the data. Therefore, to further understand the nature of individual fluctuations, the relative changes in firing rate were calculated for each neuron i as

$$\delta r_i = \frac{r_{i,t2} - r_{i,t1}}{r_{i,t2} + r_{i,t1}} \tag{2.1}$$

for common channels between consecutive recordings. In Figure 2.19 the distributions of absolute values of those relative changes are presented. In all cases, it transpires that at least half of the neurons changed their firing rate by more than 0.1, and in

most cases by more than 0.15. This corresponds to approximately 20-30% changes in the firing rate. Furthermore, for nearly all pairs of recordings a quarter of neurons changed their firing rate by more than 40%.

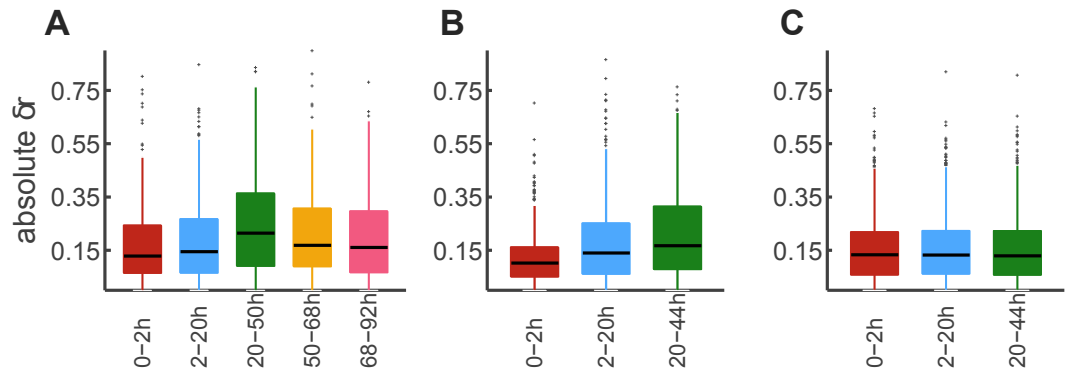

Figure 2.19: Magnitude of relative changes in firing rates between recordings. (A) The distributions of absolute relative changes in firing rates δr_i of individual units in the form of a boxplot, culture 1. (B-C) Same for cultures 2 and 3.

As a further confirmation of the significance of those fluctuations, medians of the distributions of relative firing rate changes were analysed in a manner analogous to Section 2.3.1. Namely, CIs were estimated with bootstrap; however in this instance it was in order to assess whether the medians were significantly different from null. This was indeed the case in all pairs of recordings, as presented in Figure 2.20.

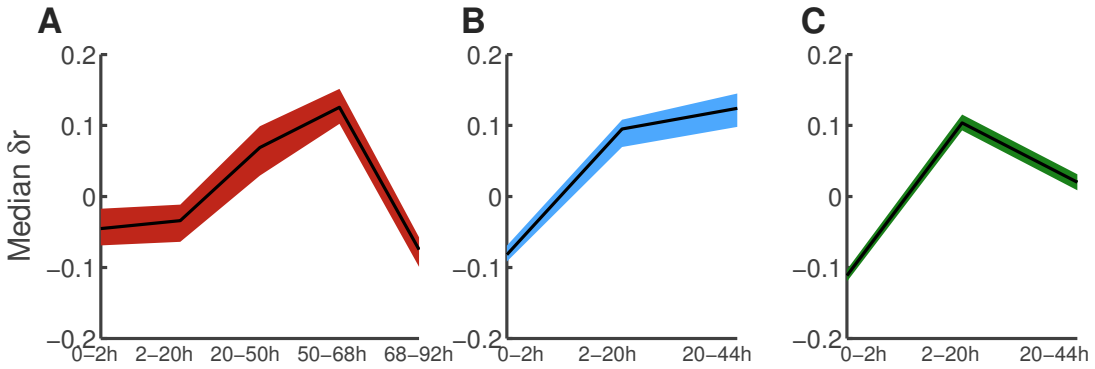

Figure 2.20: Median relative changes in firing rates between recordings. (A) The medians of distributions of relative changes in firing rate δr_i between consecutive recordings, with confidence intervals estimated by bootstrap, culture 1. (B-C) Same for cultures 2 and 3.

Finally, in Figure 2.21 the individual neurons' δr_i are plotted as a function of firing rate, uncovering a somewhat nontrivial relationship between the two. It appears that in many cases there was a tendency of high-firing neurons to decrease their activity,

and the inverse held for low-firing neurons. Since, however, the plotted changes are computed in relation to a sum of the firing rates from two recordings, it does not necessarily follow that the distributions of firing rates should get narrower due to that. And indeed this is not the case, as already shown in Section 2.3.1.

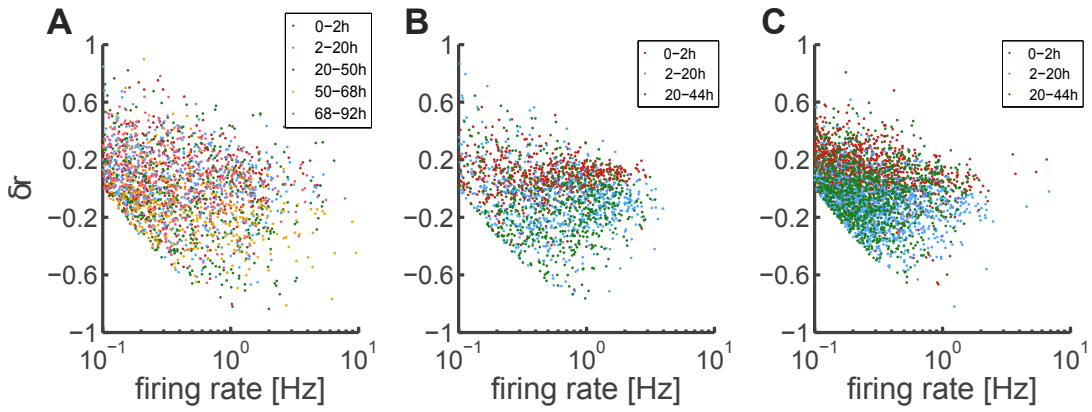

Figure 2.21: Scatter plots of relative changes in firing rate against the firing rate. (A) The relative changes in firing rate δr_i between consecutive recordings, plotted as function of firing rate in the first of the two recordings, culture 1. Log abscissa used for better clarity. (B-C) Same for cultures 2 and 3.

To further illustrate the activity fluctuations between recordings, correlations were also examined in a similar manner. However, since correlations can also take on negative values, the relative change between consecutive recordings was computed using absolute values

$$\delta c_{ij} = \frac{|c_{ij,t2}| - |c_{ij,t1}|}{|c_{ij,t2}| + |c_{ij,t1}|} \tag{2.2}$$

for each pair of neurons i and j.

As evident from examining the magnitudes (Figure 2.22), individual correlations changed relatively even more than firing rates. In both culture 1 and culture 2 more than half of the neuron pairs changed by over 50%, and in culture 3 by over 30%. This result is not surprising, given the fact that the distributions of correlations did change significantly between consecutive recordings, as established in Section 2.3.1.

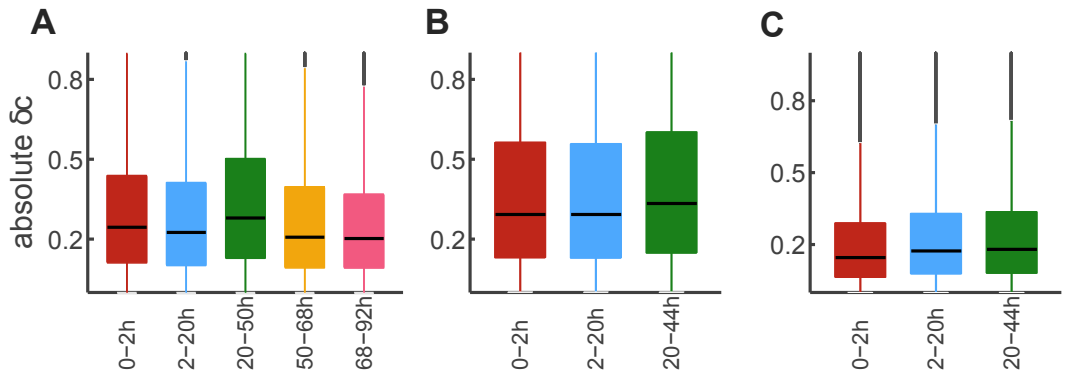

Figure 2.22: Magnitude of relative changes in correlations between recordings. (A) The distributions of absolute relative changes in correlations δc_{ij} of individual unit pairs in the form of a boxplot, culture 1. (B-C) Same for cultures 2 and 3.

However, an interesting relationship can be observed between the relative correlation change and the correlation value. As illustrated in Figure 2.23, it appears that the highest correlations tended to change the least, as indicated by the funnel shapes of the histograms (although in some cases this might be difficult to discern, since the distributions were skewed). On the other hand, in many cases there was a tendency of low-correlated neurons to increase their correlation. These results further highlight the complicated nature of the relationship between local versus global changes in the activity of neurons.

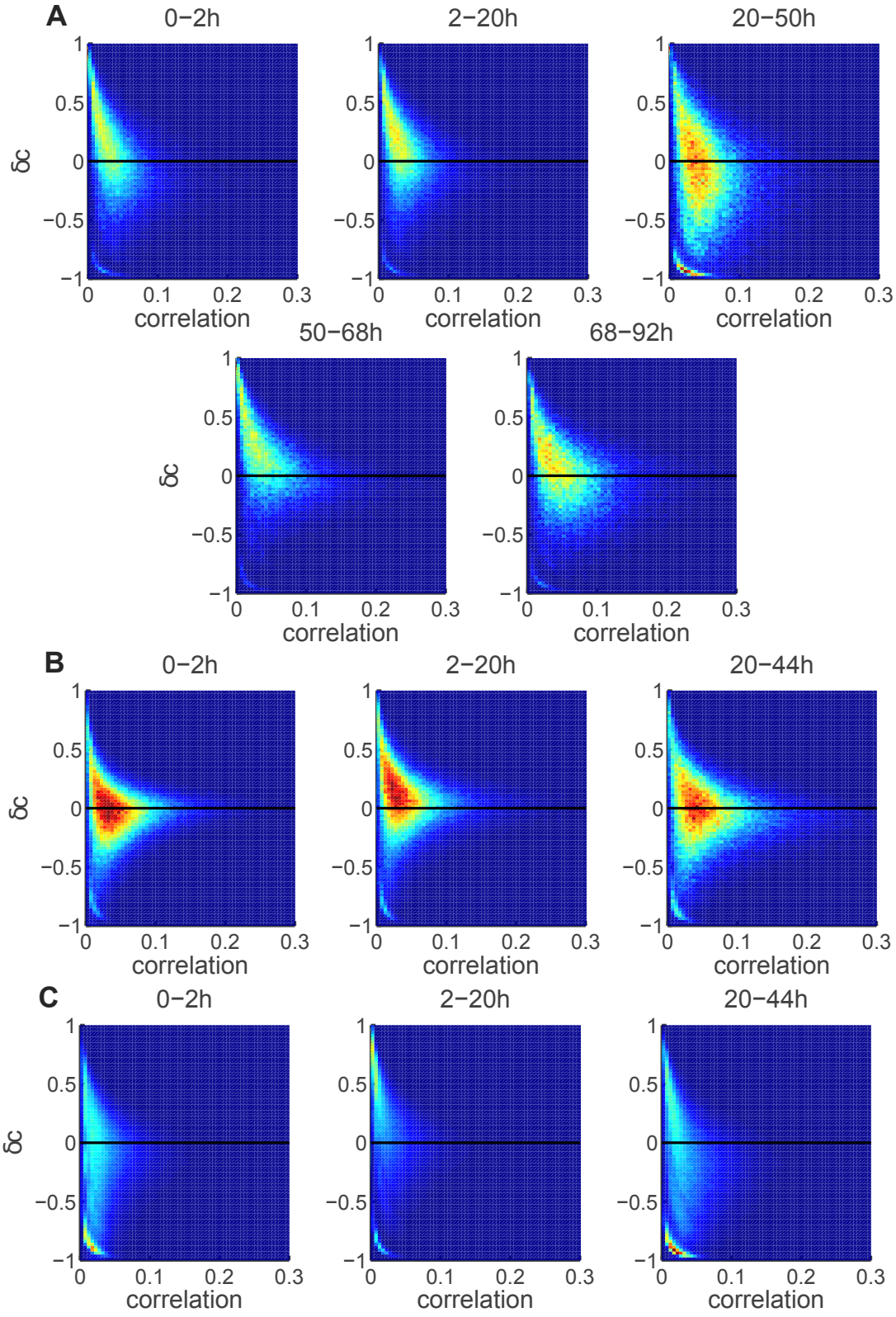

Figure 2.23: Distributions of relative correlation changes. (A) The 2D histogram of the relative changes in correlations δc_{ij} of individual unit pairs, binned along the abscissa according to the correlation coefficient in the first of the pair of recordings; culture 2, 0-2h. (B-C) Same for pairs of recordings: 2-20h and 20-44h.

2.4 DISCUSSION

In this chapter of the thesis, cultured hippocampal activity recorded with high-density MEAs is introduced as a model system for examining the interplay of global stability and local malleability of neuronal activity.

The recorded data presents hallmarks typical of such preparations, including skewed firing rates, low but significant correlations, and sparse activity with occasional highly synchronized complex bursts (Kamioka et al., 1996; Segev et al., 2001, 2004; Chiappalone et al., 2006; Wagenaar et al., 2006; Shahaf et al., 2008). Furthermore, similar to what Slomowitz et al. (2015) observed, the firing rate distributions exhibit the heavy-tail akin to log-normal distribution (although no formal test was conducted). Indeed, a recent study on the same 4096-electrode recording system in hippocampal dissociated cultures confirms this assertion (Amin et al., 2015). These observations reveal an interesting analogy to the brain, as log-normal distributions appear in many contexts *in vivo* (Margolis et al., 2012; Buzsáki and Mizuseki, 2014), and perhaps are an intrinsic feature of organization of neural circuits (Buzsáki and Mizuseki, 2014).

Dissociated cultures share many other similarities with the brain and *in vivo* activity (Marom and Shahaf, 2002) and at the same time provide superior accessibility for long-term recordings (Lütcke et al., 2013) and manipulation, which is what makes them an excellent and widely used model system (discussed by Marom and Shahaf (2002); examples include Kamioka et al. (1996); Bi and Poo (1999); Bettencourt et al. (2007); Shahaf et al. (2008) and many others). However, these preparations are randomly wired and lack any physiological structure that might be a crucial determinant of certain aspects of activity or plasticity *in vivo*. This limitation makes it difficult to generalize any findings to be applicable in the living brain. On the other hand, it is exactly this feature that allows researchers to study universal principles of self-organized circuits and abstract away from the particulars of a given brain area (Marom and Shahaf, 2002). Furthermore, since cultures are an isolated system and only participate in spontaneous activity, any changes, be it synaptogenesis, synaptic scaling, homeostatic intrinsic plasticity *et cetera*, are presumed to be ongoing natural processes. This is not as clear in *in vivo* experiments, where anaesthesia, attention, ongoing learning etc, likely modulate both activity and plasticity. For the purpose of

the present study, dissociated cultures were thus deemed superior in terms of both feasibility and universality.

Another important virtue of employing dissociated neural cultures here was the novel recording system with 4096 embedded electrodes and very high spatial resolution (Berdondini et al., 2005, 2009a,b). This allowed for simultaneous monitoring of several hundreds of single units, a wealth of data that the current study can only begin to address. One interesting direction for further development would be assessment of the proportion of 'dark neurons' (Shoham et al., 2006), i.e. healthy and responsive but not intrinsically active cells (or active at extremely low rates). As Shoham et al. (2006) point out, such units are routinely missed in *in vivo* recordings due to selection bias, and scientist are only beginning to appreciate the extent of sparsity of neural activity (Margolis et al., 2012). Multielectrode arrays and dissociated cultures provide an excellent platform to combine various recording modes and imaging to address this issue in depth. Already in the datasets presented here, the highly skewed firing rate distributions and the fact that active electrodes constitute no more than a third of available recording sites, suggest the presence of many 'dark neurons'.

The three sets of recordings were obtained from non-sister cultures, to allow for inter-preparation variability, as suggested in (Wagenaar et al., 2006). Indeed, the firing rate distributions and bursting patterns exhibited noticeable differences between preparations (e.g. Figure 2.14 and Figure 2.13), similar to earlier reports (Wagenaar et al., 2006). It should be noted however that the topologies of the networks as well as the numbers of detected active neurons also differed substantially (see Figure 2.12), which likely also contributed to the variability. In general, the influences of random factors during development and particular genetic traits of individual animals are not easy to separate, and potentially interact (Corner, 2008). However, the collected dataset appears sufficiently diverse so as to exclude the possibility that results obtained in this work are specific to a particular batch of neuronal preparations.

At the age of 24 DIV, the hippocampal cultures used here were expected to be in a mature regime of behaviour (Nakayama et al., 2005), and indeed the complexity of bursting behaviour (see Figure 2.9) conforms with that notion (Kamioka et al., 1996; Marom and Shahaf, 2002; Chiappalone et al., 2006). Moreover, this bursting behaviour remained largely stable throughout the experiments, as indicated by the fact that the distributions of interspike intervals pooled within each recording re-

mained prominently more correlated within-culture than across different cultures (see also Figure 2.17). A further indication of maturity and stability was exhibited in the behaviour of population firing rates, in particular in the fact that the skewed distributions particular to each preparation remained largely unchanged throughout the duration of the experiments (Figure 2.14).

In contrast to population stability, the firing rates of single neurons did not remain stable and exhibited a range of behaviours, from relatively small fluctuations to large, pronounced changes in firing (summarized in a scatterplot in Figure 2.21). Similarly, pairwise correlations between units changed from recording to recording, possibly reflecting functional connectivity changes. This essentially reproduces the findings of Slomowitz et al. (2015), who also found the discrepancy between global stability and local dynamics. All these results are resemblant also of the study by Minerbi et al. (2009) and their observation of stable distributions of synaptic sizes, accompanied by significant changes in individual synapses themselves. It appears thus that local properties of neurons in cultures, both structural (Minerbi et al., 2009), functional (Slomowitz et al., 2015) and activity-related (Slomowitz et al. (2015) and the current report), all undergo significant ongoing spontaneous remodelling. To exactly what extent the activity changes reflect the underlying functional and structural changes remains unclear. However, such assessment is beyond the scope of the present study, and further analyses focus solely on spiking activity.

In conclusion, the dissociated cultures recorded here constitute a suitable model system for the purpose of the study. All specimens exhibited stable global behaviour accompanied by natural fluctuations in activity of single neurons, congruent with literature and indicative of ongoing plasticity.

3

PAIRWISE MAXIMUM ENTROPY MODELLING

3.1 BACKGROUND

With the development of neuroscience and a growing volume and complexity of experimental data, analysis techniques increasingly draw from other disciplines to better understand the brain. The pairwise maximum entropy model (also referred to in the neural literature as the Ising model [1]; these two names will be used interchangeably in this work) is one such example, building on insights from statistical physics and probabilistic inference to introduce a new point of view and provide fresh insights into the nature of neural data (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008; Yu et al., 2008; Ohiorhenuan et al., 2010). It is worth noting here that it is a very particular example and a uniquely meaningful model. Although it was initially introduced in neuroscience as a purely data-driven probabilistic approach (Schneidman et al., 2006) operating on minimal prior assumptions, its functional form happens to correspond to a class of models in physics called spin glasses, and to the computational model of content-addressable memory proposed by Hopfield (1982). Thus, it emerges both as a data-driven and a hypothesis-driven model of neural activity (Schneidman et al., 2006). Furthermore, as a member of a broader class of physical models described by the Boltzmann distribution, it links thermodynamics with information theory (Crooks, 2007), a relationship that is both philosophically interesting and increasingly practically relevant in science (Parrondo et al., 2015).

1 The Ising model itself was initially invented by the physicist Wilhelm Lenz as a mathematical description of ferromagnetism, predicting the probabilities of possible arrangements of interacting magnetic spins on a lattice. It was subsequently named after his student Ernst Ising, who derived an exact solution for the 1-dimensional case of the problem (Brush, 1967). In its original form it is not equivalent to what is currently known as the Ising model in neuroscience, since it only considered nearest-neighbour interactions between spins (Brush, 1967); however, it was later generalized to include all-to-all connectivity, such as in spin glasses, or in the work of Hopfield (1982). The Hopfield model was one of the earliest adaptations of this statistical physics approach to understanding the neural system, and particularly the emergence of content-addressable memory. However, it differs from the maximum entropy model in the fact that it is not a generative model fit to stationary distributions of neural data, but rather an abstracted model with a stochastic time-based update rule, allowing for examining the evolution of the system in phase space.

The model was first adopted by neuroscientists in order to better understand the population code of the retina (Schneidman et al., 2006). In essence, when examining the occurrence of concurrent spiking of two or more neurons, the authors found that those appear – on average – statistically much more often than coincidental firing predicted from the rates under assumption of independence (see Fig.2 a) and b) of Schneidman et al. (2006) and also Figure 3.4 A and B of the present work). However, typical correlation values of neurons in the retina – and in fact of neurons in general in many types of recordings – are low, lending little explanation to the appearance of multi-neuron spiking patterns. The authors attempted to tackle that discrepancy by directly modelling the distribution of the occurrence of spiking patterns in a group of neurons as a minimally structured distribution under the constraints of mean firing rates and correlations, and by only assuming pairwise interactions between neurons. They did so for multiple random sub-groups of 10 neurons each, since this size was small enough for exact fits and proper statistical evaluation, but large enough to reveal departures from independence. Remarkably, this model – the Ising model – was shown to capture 'more than 90% of the structure in the detailed patterns of spikes and silence in the network' (Schneidman et al., 2006). This assertion was based on the fact that the employed model performed orders of magnitude better than predictions from the rates and was shown to explain over 90% of the entropy difference between data and independent-prediction pattern distributions in the retina, and over 95% in a cortical culture. Importantly, the fit of a pairwise maximum entropy model was more than simply a method of dimensionality reduction. Firstly, the fact that this particular form of model distribution provided a very good fit to the data, i.e. that pairwise interactions were sufficient to explain multi-neuron correlation structure, suggests that the ganglion cells of the retina might be operating under the principles of an associative neural network (Hopfield, 1982) and posses error-correcting properties (Schneidman et al., 2006). Secondly, the parameters of the Ising model are easily interpretable, which brought promise of a deeper understanding of functional connectivity than correlations could offer. For example, Schneidman et al. (2006) illustrated how 'frustration' in an Ising model (i.e. triplets of units with an odd number of negative interactions) results in spurious correlations. Finally, from analysis of groups of increasing size the authors predicted a 'critical' network size, which appeared to agree with physiological data (however, this particular result has been challenged by

further research (Roudi et al., 2009a), which questioned the validity of extrapolation performed by Schneidman et al. (2006)).

More work followed with other researchers extending the above results. In another study on retina Shlens et al. (2006) employed a simplified version of the pairwise maximum entropy model, using only adjacent cell interactions and thus further reducing the number of parameters used. Not only did the model perform well, it in fact surpassed the results of Schneidman et al. (2006), as it was shown to account for over 98% of departures from independence (although in smaller groups of neurons and with a different applied visual stimulus than the original study, both of which could have influenced this favourable difference). Authors also analysed spatially extended groups of neurons to demonstrate that even long-range correlations can be explained by a series of adjacent-cell interactions. Thus Shlens et al. (2006) have further shown that often quite complicated and globally synchronized states can be attributed to relatively simple – but numerous – pairwise interactions.

Yu et al. (2008) have taken this area of research further still, employing the Ising model for *in vivo* recordings from visual cortex of the cat, and then building on the results of modelling to infer the underlying properties of networks of neurons. Similarly to Schneidman et al. (2006), the authors looked at many randomly chosen groups of 10 neurons each. After establishing adequate fit quality (92-93% explained entropy difference) authors then focused on the inferred interactions. Since they extensively sampled the available population, they were able to evaluate each interaction parameter between two individual neurons on average 50 times, due to both neurons belonging to an average of 50 groups. The variability of the parameter for individual pairs across groups was much smaller than that across different pairs of neurons (standard deviation of 0.008 as opposed to 0.12), allowing the authors to conclude that the fits gave a reliable estimate of functional connectivity. They then proceeded with topological analysis of all 10-neuron groups and of the reconstructed (by averaging over groups) complete population group and found that the networks had small-world properties. Interestingly, the small-worldness was functionally relevant, with strong clustering observed in units with similar orientation preferences, and a low number of connections bridging different cell types. Thus in the study of Yu et al. (2008) the Ising model fulfilled its promise of novel insight due to its interpretable parameters.

Further exploration of the applicability but also of the limitations of the model came from the work of Tang et al. (2008). First, the authors examined in detail the fits of an Ising model to different kinds of neural data (Tang et al., 2008), including hippocampal and cortical cultured rat neurons, acute rat cortex slices, and acute human cortex slices. The results largely replicated earlier observations, although the performance of the model varied across preparations, with the lowest score of approx. 82% in dissociated cortical cultures. Second, Tang et al. (2008) sought to examine whether the statistical predictions of the model would extend to temporal domain. To test that, authors drew network states randomly from the modelled distribution, and then compared the lengths and sizes of concatenated temporally adjacent firing patterns to the ones present in the recorded data. Unsurprisingly, the model failed to reproduce the data-observed spatio-temporal patterns, confirming that the neural data possesses nontrivial temporal structure. The above result was not surprising because the Ising model in the above example was fit to data from single time bins, and the second order marginals that it was meant to reproduce reflected only spatial, but not temporal correlations. This automatically results in lack of explanatory power when it comes to sequences of complex spiking patterns, such as in the case of the above-mentioned failure observed by Tang et al. (2008), the bursts discussed in the previous Chapter 2, or neuronal avalanches (Yu et al., 2011).

Another limitation of the pairwise maximum entropy approach that did not present issues in the initial study (Schneidman et al., 2006) is its lack of higher-order interaction terms. While it appeals to intuition to model networks of neurons as operating solely on the basis of pairwise connections (which are conceptually mappable to synapses), it transpires from *in vivo* research that at times the pairwise model is not sufficient (Ohiorhenuan et al., 2010). The authors recorded simultaneous spike trains from tetrodes implanted in primary visual cortex of *Macaca mulatta* and, as in other studies, fit the Ising model to the distributions of patterns from multiple groups of a few neurons each (3-6). The groups were chosen in a distance dependent fashion, with neurons clustering either within a 150μm radius, or separated by 600μm, or by more than 1000μum. This way the authors were aiming to distinguish between within-column and longer-range connectivity. As transpired from this analysis, despite being orders of magnitude better than the independent model, the Ising model often failed in the short range (as assessed by comparing the Kullback-Leibler divergence be-

tween model and data to that of data-to-data divergence). Such failures are indicative of a scale-dependent difference in the importance of higher-order interactions in neural activity. Furthermore, this effect was modulated by stimulus conditioning, which suggests that the prevalence of beyond-pairwise interactions was state-dependant. However, it is not readily testable whether the detectability of higher-order terms indicated the presence of nonlinearities, special type of multi-neuron coupling, or simply a strongly connected unobserved neuron. The implications of the above distinction are important for further research; if it is a matter of multi-neuron coupling or nonlinearity, alternative models to the Ising one should be considered. If, on the other hand, it is a question of including in inference all relevant units, it might be that the pairwise maximum model is sufficient, only it fails when presented with insufficient information.

In another insightful study, Yu et al. (2011) explicitly tackled the above issue by introducing a novel model, the Dichotomised Gaussian (Macke et al., 2009), and benchmarking it against the pairwise maximum entropy. The Dichotomised Gaussian is a near-maximum entropy model that is fit to the same set of marginals as Ising, but which, owing to a nonlinearity introduced by thresholding, can accommodate higher-order correlations without introducing additional parameters. The data came from *in vivo* recordings from the macaque (as in Ohiorhenuan et al. (2010)), and the authors used 10-neuron groups (similarly to Schneidman et al. (2006) and Tang et al. (2008)), both of which allowed for useful comparisons against the existing literature. Furthermore, Yu et al. (2011) applied the model not only to spiking activity, but also to patterns of thresholded local field potentials (LFPs) and to neuronal avalanches (i.e. patterns of thresholded LFPs where the active and neighbouring time bins were concatenated). Several important results transpired from their analysis. Firstly, a substantial portion of higher-order correlations in neuronal avalanches was due to temporal correlations, and while the Ising model was not able to adequately fit these (mirroring the results of Tang et al. (2008)), the Dichotomised Gaussian was. Secondly, the statistics of non-concatenated LFPs also exhibited higher-order effects and again the novel model outperformed the maximum pairwise entropy. Finally, the authors also found that their model performed better than Ising when applied to spiking activity, although in this case the improvement was less pronounced than in the case of local field potentials. The authors suggested that in the latter two cases the higher-

order correlations were likely attributable to the thresholding nonlinearity and then further confirmed the viability of this hypothesis with simulations. Importantly, in these simulations Yu et al. (2011) were also able to tentatively quantify the amount of higher-order effects under varying parameter regimes, at the same time uncovering a relationship between higher-order terms and first and second order marginals. The authors concluded that beyond-pairwise correlations are most abundant when the firing rates are low but the correlations are large. Thus Yu et al. (2011) were able to dissociate and tentatively pin-point separate sources of higher-order effects and propose an explanation as to the varying performance of the Ising model across studies. Interestingly, their results do not invalidate the broader conclusions of earlier research, whereby complex statistics of activity can be explained by numerous pairwise connections, since what transpired from their research was that most of higher-order effects were simply a by-product of a thresholding nonlinearity.

In summary, the pairwise maximum entropy model has proven to be a useful tool for neuroscience, providing important insights into the synchronous activity of neurons both through its successes (Schneidman et al., 2006; Shlens et al., 2006; Yu et al., 2008) and failures Ohiorhenuan et al. (2010); Yu et al. (2011). Specifically, it has demonstrated the viability of significant and meaningful dimensionality reduction in the description of ensembles of cells. This result is of particular import here, since the focus of this work is to better understand emergent stability in populations of neurons and the mapping provided by the Ising model between individual parameters and group activity allows for a deeper probing of these phenomena than simple analysis of rates and correlations. Of course, as in the case of any model, it needs to be taken with caution, given its known limitations (Tang et al., 2008; Ohiorhenuan et al., 2010; Yu et al., 2011). However, since the hippocampal cultures employed in this work are characterized by both low rates and low correlations (favourable to the Ising model, as found by Yu et al. (2011)), and do not possess the fine columnar structure of the cortex (which could have been the reason for the model failures in Ohiorhenuan et al. (2010)), and given its parsimonious and analytically tractable description of the data, it was considered the optimal choice for the purpose of this study.

3.2 METHODS

The analyses described here and in the next chapter were all implemented in Matlab by the author. The complete code is made publicly available as a Dropbox folder at the following web address: https://www.dropbox.com/sh/v2qcfeerxoorjdk/AAD1gKK-Uz_3HOywsJ-XYL5Ma?dl=0

Aside from the code and documentation, the package includes as well the original data used in this work, to allow for independent reproducing of the results, as well as for any independent follow-up investigations. A final approved copy of the thesis will also be included in the folder for technical reference on the methods.

Should for any reason Dropbox services become unavailable or obsolete in the future, all of the above can be also obtained through e-mailing the author at:

dagmara.panas@gmail.com

3.2.1 Theoretical foundations

In the most general terms, maximum entropy modelling is a type of statistical inference, i.e. a means of approximating probability distributions from imperfect, incomplete data. Performing such an approximation carries two possible advantages: overcoming prohibitive dataset sizes, where exact calculation of the probability distribution is impossible; and obtaining an analytical form of said probability distribution, which in certain cases (as here) can facilitate further analyses of the system. The core idea of the method is to find a probability distribution that matches the observed statistical regularities of the data and is otherwise as *un*structured as possible. In mathematical terms, this translates to the distribution having maximal attainable entropy (hence the name of the method) given the specified statistical constraints, where entropy is defined as follows

$$S_p = -\sum_X p(X) \ln\left(p(X)\right) \tag{3.1}$$

and can be commonly understood as a measure of disorder. Therefore, the above requirement can be interpreted simply as imposing the minimal necessary assump-

tions on the structure of inferred probability distribution, an approach conceptually equivalent to Ockham's razor.

In principle, any number of measurable statistics defined by f_c (c being here an index of employed statistics) can be used as constraints placed upon the sought distribution p over the possible states of the system

$$\sum_{\{X_s\}} p(X_s) f_c(X_s) = E\left(f_c\right)_{data} \tag{3.2}$$

with X_s being a vector of length N representing the state of an N-element system and $E(.)_{data}$ denoting the expected value under the data distribution, i.e. the observed statistic. With the additional requirement of the entropy being maximal, and the necessary assumption of p being a probability (i.e. summing to unity), this leads to the following equation specifying the model (Cover and Thomas, 2012)

$$p = \arg\max_{p_d} \left[-\sum_{X_s} p_d(X_s) \ln\left(p_d(X_s)\right) + \sum_c \Lambda_c \left(\sum_{X_s} p_d(X_s) f_c(X_s) - E\left(f_c\right)_{data} \right) \right] \tag{3.3}$$

where Λ_c are the Lagrange multipliers (including λ_0, for which f_0 is unity and $E(f_0) = 1$, thus ensuring that the probability sums to 1).

Solving for p yields a closed-form solution (Cover and Thomas, 2012)

$$p(X_s) = \frac{\exp(-\sum_c \Lambda_c f_c(X_s))}{\sum_{X_s} \exp(-\sum_c \Lambda_c f_c(X_s))} \tag{3.4}$$

with the free parameters Λ_c ensuring that conditions specified in Equation 3.2 are met.

This functional form resembles the Boltzmann distribution (also called the Gibbs distribution) in statistical physics

$$p(i) = \frac{\exp(-E_i/kT)}{\sum_i \exp(-\sum_i E_i/kT)} \tag{3.5}$$

where $p(i)$ is the probability of finding the physical system in state i, E_i denotes the energy of that state, k is the Boltzmann's constant, and T is the temperature of the system. The similarity between the Boltzmann distribution and maximum entropy model solution has interesting consequences: if the statistical constraints in maximum entropy modelling are chosen appropriately, the sought distribution is exactly the

Boltzmann distribution, and the expression under the exponent can be understood as the energy of the system. In those cases, the fitted Λ_c acquire a physical interpretation, rather than being just abstract parameters.

This is the case for the pairwise maximum entropy model considered in this work. The constraints placed on the distribution are the first and second order statistics of the data

$$f_1 = X_s \tag{3.6}$$

$$f_2 = X_s X_u \tag{3.7}$$

and under such constraints Equation 3.4 becomes

$$P(X_s) = \frac{\exp(-\sum_j \lambda_j x_j - \frac{1}{2}\sum_{k\neq j} \lambda_{jk} x_j x_k)}{\sum_{\{X_s\}} \exp(-\sum_j \lambda_j x_j - \frac{1}{2}\sum_{k\neq j} \lambda_{jk} x_j x_k)} \tag{3.8}$$

(where λ_j are elements of Λ_1 and λ_{jk} are elements of Λ_2). This distribution is equivalent to an infinite-range Ising model, i.e. a model describing a lattice of magnetic spins interacting with each other and with an external magnetic field. Under this analogy, the Lagrange multipliers λ_j and λ_{jk} can be understood as individual magnetic field acting on unit j and magnetic interaction strength between units j and k, respectively.

It is exactly this analogy that makes the pairwise maximum entropy model particularly apt for neural data, because the individual field can be conceptually understood as effective excitability of a neuron, and the magnetic interaction as functional connectivity between two neurons. The concept of effective excitability and functional connectivity as sole determinants of multi-neuronal firing has been introduced in early models of artificial neural networks, namely the Hopfield network (Hopfield, 1982). This has later been extended to include stochasticity, and is known as the Boltzmann machine (Ackley et al., 1985), which in a steady state is specified by the very same equation that specifies the pairwise maximum entropy model (Equation 4.5). Therefore, the Ising model fitting can be also understood as the inverse problem of the Boltzmann machine. Thus, the modelling method chosen for this work is more than simply a means to obtain a functional form of the probability distribution of network states, it is also a fit of a simple but viable spiking neuronal network model.

3.2.2 Model fitting procedure

Fitting a pairwise maximum entropy model to recorded spiking data requires finding appropriate parameters λ_j and λ_{jk}. Since the problem is well-specified (no local maxima exist since the log-likelihood is convex with respect to parameters, as can be seen in Section 4.2.2) any gradient ascent optimization method can be used for that purpose. Here, the procedure was as follows. First, spike trains of a group of N neurons were binned, creating binary matrices $X = \{x_j^t\}$ with +1 representing a spike and -1 silence of unit j in bin t. In each time bin, the state X^t of the group could be one out of 2^N possible spiking patterns X_s (see Figure 3.1 for an illustration for the case of 4 neurons). The fraction of how often state X_s occurred in the data yielded the observed probability distribution of spiking patterns in the group $P_{obs}(X_s)$ – the distribution to be modelled. This distribution will typically depend on the choice of a time bin, as larger time bins are more likely to include more simultaneous spikes and thus exhibit more synchrony, possibly including spurious higher-order correlations. On the other hand, time bins that are too small will ignore existing and pertinent synchrony. As discussed in Section 2.2.6, the particular choice of 5ms made in this work was dictated by consideration of the time scale of latency of a monosynaptic connection.

Then, the constraints for the pairwise model were calculated

$$\langle x_j \rangle_{obs} = \frac{1}{T} \sum_{t=1}^{T} x_j^t \tag{3.9}$$

$$\langle x_j x_k \rangle_{obs} = \frac{1}{T} \sum_{t=1}^{T} x_j^t x_k^t \tag{3.10}$$

and to obtain a model fit the parameters were adjusted by an iterative scaling algorithm, as in Tang et al. (2008)

$$\delta\lambda_j = \eta \times \text{sign}\left(\langle x_j \rangle_{obs}\right) \times \ln\left(\left|\frac{\langle x_j \rangle_{obs}}{\langle x_j \rangle_{mod}}\right|\right) \tag{3.11}$$

$$\delta\lambda_{jk} = \eta \times \text{sign}\left(\langle x_j x_k \rangle_{obs}\right) \times \ln\left(\left|\frac{\langle x_j x_k \rangle_{obs}}{\langle x_j x_k \rangle_{mod}}\right|\right) \tag{3.12}$$

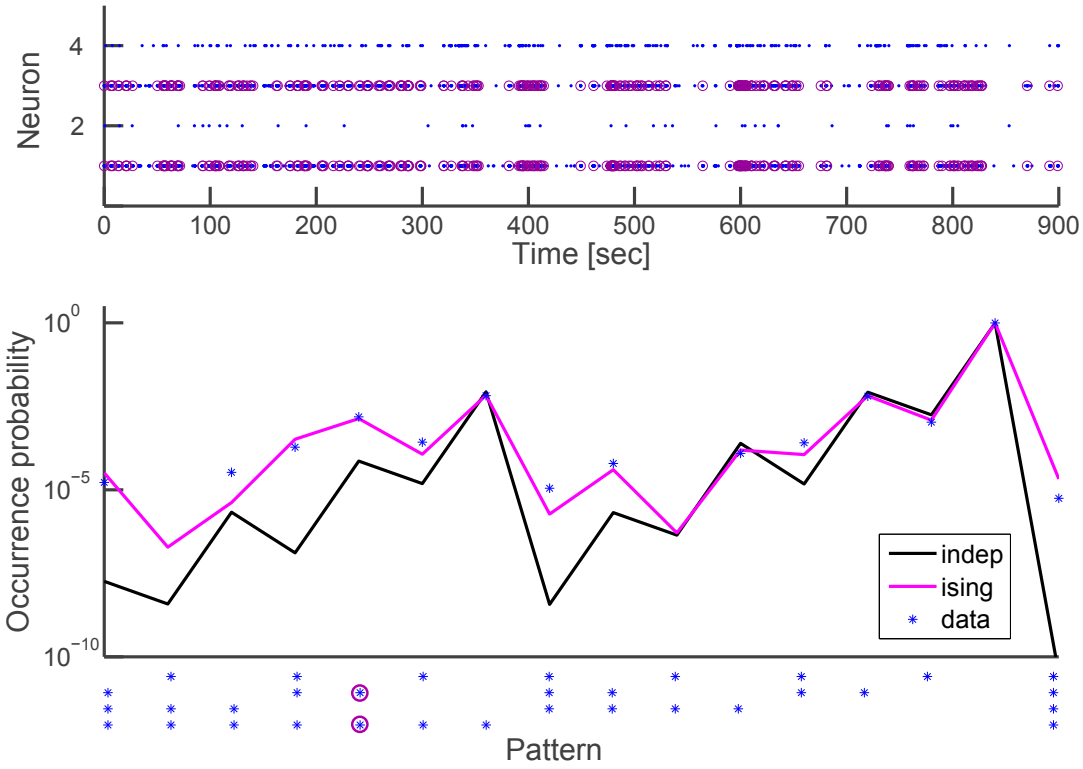

Figure 3.1: Illustration of the pairwise maximum entropy modelling procedure for neural data. Top panel: a raster plot of spiking activity in a randomly chosen group of 4 neurons; circled in magenta are the instances of a particular firing pattern [1 -1 1 -1] – the number of those instances divided by the number of time bins in the recording is the observed probability of that pattern. Bottom panel: the observed probability distribution of all possible spiking patterns in the same group of neurons (blue dots) – this is the probability distribution to be modelled; pairwise maximum entropy model distribution (pink) as an approximation of the observed distribution, in comparison to the independent model (black).

The model observables were re-calculated at each iteration, as the $P(X_s)$ (Equation 4.5) was updated

$$\langle x_j \rangle_{mod} = \sum_{X_s} P(X_s) x_j \tag{3.13}$$

$$\langle x_j x_k \rangle_{mod} = \sum_{X_s} P(X_s) x_j x_k. \tag{3.14}$$

A constant learning rate $\eta < 1$ was used for each set of neurons (initially a value of 0.75 was chosen, following the literature (Tang et al., 2008); however it was found in tests that often even a higher learning rate could be adopted, speeding up the computations; therefore typically $\eta = 0.9$ was used). Adjustments continued until the relative difference between model and observed marginals was no larger than

10^{-5}, which corresponds to reproducing spike and coincident spike numbers with the precision of approximately 1 spike. If a group did not converge within a prescribed number of iterations (typically 60000), alternative value of η was used, to obtain the desired convergence (spanning a range of $0.1 - 0.6$). In the rare cases of incomplete convergence after testing various learning rates, the precision dropped to approximately 1.5-2 spikes.

3.2.3 Choice of neuron groups

The fitting procedure described above requires calculating the modelled probability distribution at each iteration step. Since there are 2^N possible configurations of an N-neuron group, this process fast becomes too computationally intensive to be viable. It is certainly not feasible to fit an Ising model to the entire population of neurons from a high-resolution *in vitro* recording, even employing computational approximations such as Monte Carlo sampling or analytical approximations such as those described by Roudi et al. (2009b) (the group sizes that these techniques were tested with are still below the numbers of recorded units in the present study, i.e. 200 neurons in the case of Roudi et al. (2009b) versus over a 1000 units in culture). Furthermore, such fits would have been unverifiable as to their quality (Roudi et al., 2009a), since with more and larger patterns the observed probability distributions become technically more and more undersampled. Therefore, the approach here was to sub-sample the available population extensively, similarly to related work. And, since it is not clear *a priori* what constitutes a reasonable group size for the present data, therefore as in other related work (Schneidman et al., 2006; Shlens et al., 2006; Yu et al., 2008; Tang et al., 2008; Yu et al., 2011), the approach here was to sub-sample the available population extensively. The group size was chosen to be 10, following the majority of studies in this area. As Schneidman et al. (2006) pointed out, this scale is large enough to observe complex collective behaviour, while still being small enough to allow direct validation of models against observed probability distributions. Furthermore, Yu et al. (2008) observed that for random groups of 10 neurons not only was the pairwise model sufficient to explain correlated activity, it also provided consistent estimates of functional connectivity across groups.

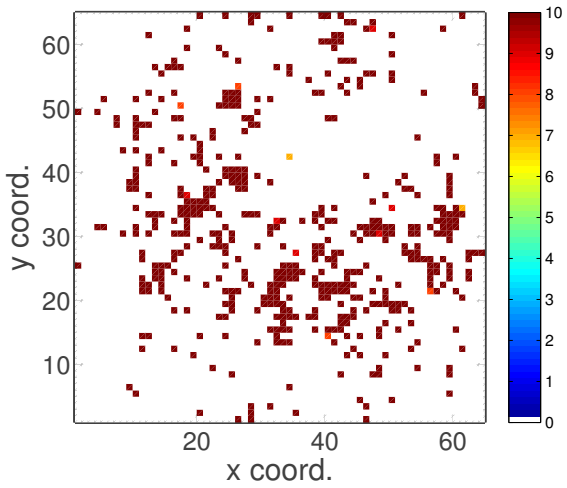

Figure 3.2: Illustration of the principle behind the sub-sampling procedure: arranged on a 64×64 grid corresponding to the electrode layout is the count of how many times each of the available active channels was chosen for a random group of neurons.

Unlike in the majority of related studies, the number of sampled groups was not fixed. Instead, the fixed value was the number of times each channel was chosen to be in a group. The aim of this modification was twofold; first, to minimise the influence of sampling bias in the regime where analysis of all possible permutations of subgroups was infeasible and the population was quite heterogeneous; second, to allow more readily for comparing different preparations with different numbers of active channels.

First, the common active channels between each two consecutive recordings were identified, and groups were then sampled from that pool of channels. This way, the model could be fit to the same group at two different time points, and then both group behaviour and individual neuron parameters could be compared across time. The 10-neuron groups were chosen at random from the available channels repeatedly, using the Matlab `randperm` function. Any channel that has already been chosen 10 times was removed from the available pool. The random sampling continued until the distribution of how many times a channel was chosen was homogeneous (for a representative example see Figure 3.2). This resulted in approx. 400-1200 groups, depending on the preparation.

3.2.4 Model evaluation

The Ising model fits converged to the desired criterion in the majority of cases, but that provides little information on how good the models were in truly explaining the data. A maximum entropy solution for given constraints is guaranteed to exist, regardless of whether the underlying probability distribution is poorly- or well-approximated. And although it was not the intention of the study to prove the Boltzmann machine model entirely sufficient in explaining multi-neuron firing phenomena, it was crucial to perform tests of fit quality for a correct interpretation of further results and cross-validation with other studies. Therefore, in order to assess how well the pairwise maximum entropy model represented the activity of groups of neurons, two measures were used: the multi-information ratio (Schneidman et al., 2006; Tang et al., 2008; Yu et al., 2008); and the Shannon-Jensen divergence (Schneidman et al., 2006; Shlens et al., 2006; Yu et al., 2011).

The first of these measures quantifies to what extent the model can account for correlated states in the network. To understand how this measure is constructed, let us consider maximum entropy models of all possible orders. The lowest, first order maximum entropy model can explain only the firing rates of individual units and is equivalent to an independent Poisson model of spiking (here referred to throughout as the independent model). Next is the second order, i.e. the Ising model, and then follow higher-order maximum entropy models with combinatorially increasing numbers of constraints. Each of these provides an increasingly better approximation of the data probability distribution, as more parameters are added. At the same time, as the distributions reflect increasing level of interactions, the disorder in the distributions decreases, and so does the entropy (Schneidman et al., 2003, 2006). Therefore, it has been proposed to measure the compound network correlation as the difference between the entropy of the first-order model and the entropy of the observed data, a.k.a multi-information (Schneidman et al., 2003, 2006). Building on that definition, the multi-information ratio is a relative measure that expresses what proportion of the total multi-information can be accounted for by the evaluated model (here, the Ising model). It is expressed as follows

$$MIR = \frac{S_{indep} - S_{ising}}{S_{indep} - S_{data}}, \tag{3.15}$$

where S stands for entropy (see Equation 3.1). In other words, the multi-information ratio tells what fraction of the observed order in the data probability distribution can be explained by factoring in pairwise, but not higher-order, connections. The multi-information ratio (hereafter referred to as MIR) is reported here interchangeably as a fraction or as a percentage, similarly to other reports in the field (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008; Yu et al., 2008, 2011).

It needs to be noted, however, that calculating entropy from limited data carries the disadvantage of introducing error, both in random and systematic form (Macke et al. (2013); see also Appendix A for a more detailed explanation). Therefore, the final value of any entropy-based measure will reflect not only the quality of the model fit, or how appropriate the model choice is, but also the above-mentioned errors. Here, on one hand the systematic errors were predicted to be small on average (Macke et al., 2013), and on the other hand the random variability was *a priori* unknown but potentially dominant. Therefore, instead of applying approximate correction terms for entropy, an alternative approach was adopted: random re-sampling from the modelled distributions. For each set of neurons from baseline recording in culture 1, 30 artificial 10-neuron spike trains were generated, of the same length as the data. These were then treated as data, i.e. the observed probability distributions were calculated, and then maximum entropy models were fit to match the appropriate pair-wise constraints. Then, multi-information ratio was computed for the re-fit models in reference to the re-sampled data. This could then be used to asses the expected range of deviations of models from data if the data were truly sampled from a maximum entropy model. This approach simultaneously takes care of both systematic and random errors via simulating their effects on the artificial data. An extended discussion of the problem and supporting results are presented inAppendix A.

However, the multi-information measure by itself is not capable of a comprehensive assessment of the success of the model, since it is defined in relation to an independent model and does not take into account the absolute deviation of either of the models from the actual data. Furthermore, it is based on comparing the entropies of distributions, rather than distributions themselves. Therefore, as a complementary measure for more direct and detailed comparisons the Jensen-Shannon (also referred

to henceforth as JSD) divergence was used. The JSD is a popular distance measure for probability distributions, the square root of which has the property of a metric

$$D_{JS}(P_1||P_2) \quad =$$

$$= \frac{1}{2}\sum_{\{X\}} P_1(X)\log_2\left(\frac{2P_1(X)}{P_1(X)+P_2(X)}\right) + \frac{1}{2}\sum_{\{X\}} P_2(X)\log_2\left(\frac{2P_2(X)}{P_1(X)+P_2(X)}\right). \quad (3.16)$$

This was used to calculate the divergence of the modelled distribution from the observed one. In order to allow a further understanding of fit quality, the JSD was also computed to compare the observed probability distributions derived from halves of the data. In this manner, the models' prediction of the data could be compared to how well predictable the data was, given the fact that the finite sampling problem, other random factors and possible nonstationarities limit the accuracy with which the data can be known.

3.2.5 Model comparison measures

The JSD was also employed for another purpose – to compare models between time points. This constituted a measure of similarity of the combined behaviour of the multi-neuron groups chosen for analysis. Keeping in mind that the goal of the study was to assess the relationship between group stability and stability in individual parameters, what was also needed was a combined score of parameter similarity. For that purpose the coefficient of determination r^2 was used.

In general, this measure is used in statistics to indicate how well a chosen model fits the data. In order to adapt it for the purpose of comparing two sets of parameters at different time points, linear regression was performed between the two groups of values. First, the N field parameters λ_j and ${}^1/_2N(N-1)$ interaction parameters λ_{jk} of each group i were pooled together to form a vector Λ_i of size $M = {}^1/_2N(N+1)$. Then, for a set of parameters ${\Lambda_i}^{t1} = \left\{{\lambda_m}^{t1}\right\}$ at time t_1 and ${\Lambda_i}^{t2} = \left\{{\lambda_m}^{t2}\right\}$ at time t_2, linear fit was carried out as follows

$${\Lambda_i}^{t_2} = a{\Lambda_i}^{t_1} + b. \quad (3.17)$$

The r^2 of this fit was then used as a measure of similarity between parameter groups

$$r^2 = 1 - \frac{\sum_{m=1}^{M}(\lambda_m{}^{t_2} - a \cdot \lambda_m{}^{t_1} - b)^2}{\sum_{m=1}^{M}(\lambda_m{}^{t_2} - \bar{\lambda}^{t_2})^2}. \tag{3.18}$$

The r^2 of a linear fit was chosen because this measure is not affected by coordinated changes in all parameters, and rather reflects the extent and magnitude of individual changes occurring within the group.

3.3 RESULTS

The maximum entropy model fits were performed as in previous work (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008; Yu et al., 2008); the only modification being the choice of neuronal groups such that they could be compared between time points. Thus, for each pair of consecutive recordings, the Ising model was fit to multiple randomly chosen groups, yielding parameter descriptions of group activity for the first and second recording. An example of such fit at one time-point is presented in Figure 3.3. The parameters of the model shown in panels A and B were determined through an iterative algorithm aiming to match the first- and second-order marginals of observed spiking activity. However, as illustrated in panel C, the resulting probability distribution of observing each possible spiking pattern was also well represented by the model. The fact that the parametric description provided by the Ising model goes beyond reproducing firing rates and correlations, and is capable of capturing multi-neuron statistics of activity, was the key finding from previous reports (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008; Yu et al., 2008) and the motivation for the present work.

Although the goal here was to build on the success of reported research, first however a verification of the reproducibility of earlier work was in order, since the data used here was novel in terms of its high resolution, firing rates profile and time binning. Therefore the results reported here are divided into three sub-sections. In the first one model fits are evaluated and compared in detail with the relevant literature. In the second sub-section, properties of model parameters are explored in order to gain a better understanding of the interpretability of the model. Finally, in the third

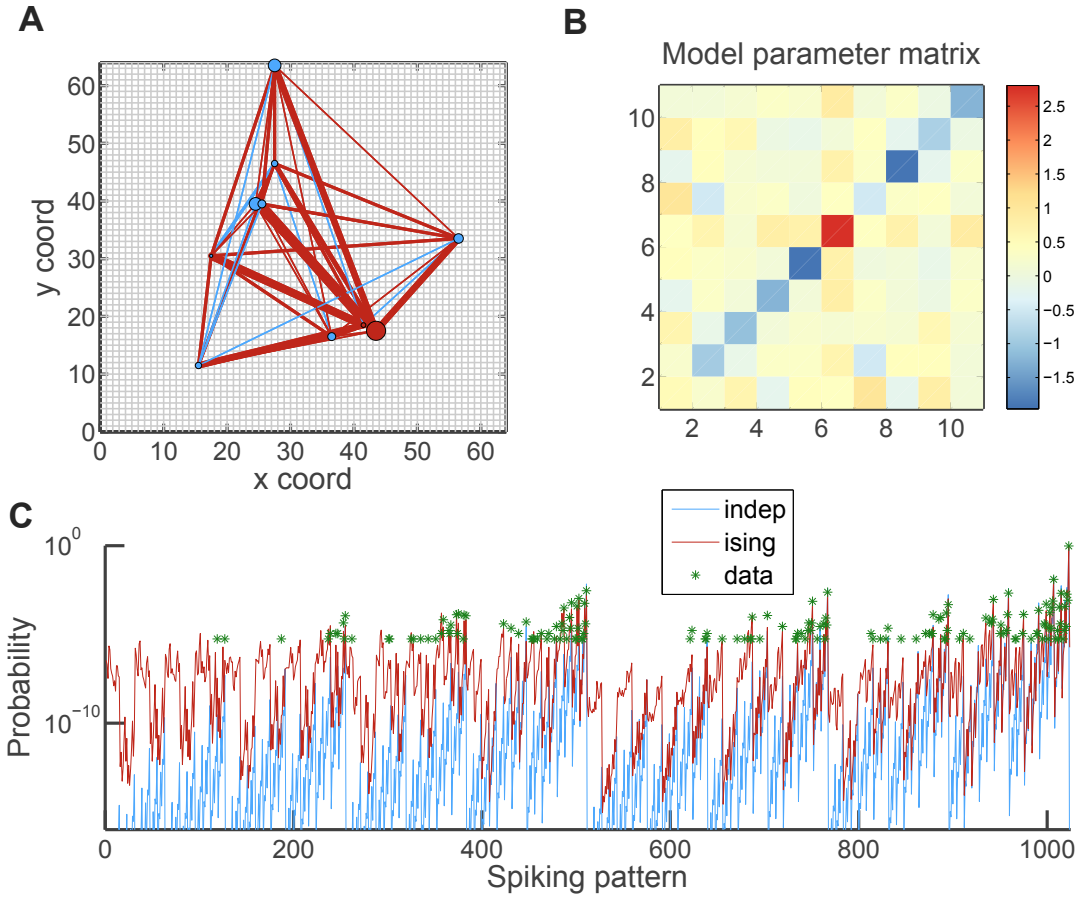

Figure 3.3: Example model fit for a random group of neurons from culture 1. (A) Parameters of the fitted model of a single group of 10 neurons at 0h (i.e. the baseline recording), pictured over the MEA layout; line widths and circle diameters are proportional to corresponding parameter values (interactions and fields); blue indicates a negative sign and maroon a positive one. (B) Parameters of the fitted model of the same group presented in the form of a matrix, with color-coded parameter values (this visualisation will become useful in the next chapter); fields are plotted on the diagonal, interactions off-diagonal; the matrix is symmetric because the interactions in the model are assumed to be symmetric; order of neurons in a group is arbitrary – a comparison with panel A shows that the 'strongest' unit (44,19) is 6^th in that group. (C) The resulting pattern probability distribution (maroon) is plotted, along with the observed one and independent one for reference; the Ising model provides a much better match for the data, predicting correctly not only the first and second order marginals, but also other multi-neuron spiking pattern occurrences.

sub-section the potential to compare models across time is explored, to illustrate the need for further analyses, warranting a separate chapter.

3.3.1 Evaluation of the maximum entropy model fits

First, to illustrate the benefits of maximum entropy modelling, in Figure 3.4A the fitted distributions are plotted against the observed ones, with the independent models pictured as a reference. For a wide range of values the model fairly represents the data (note the log-log scale), funnelling downwards only in the regime of very low probabilities. In this range, the modelled probabilities are so low as to be virtually undetectable, due to the limited length of recording (or, in other words, the confidence interval of the data-estimated pattern probability widens with decreasing rate). In panel B of Figure 3.4 as an alternative point of view the probabilities of observing a given pattern size are shown for both models and the data. Again, the Ising model clearly outperforms the independent one and appears to provide a close match of the data. Thus at first glance, the maximum entropy model appears to retain a large portion of the detailed statistical structure of the data.

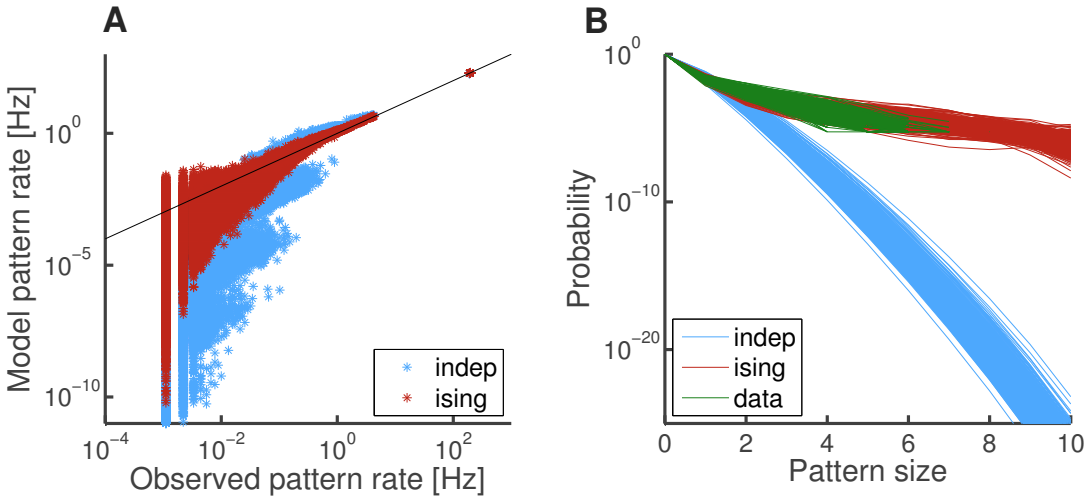

Figure 3.4: Illustration of the predictive power of the maximum entropy model. (A) The modelled probabilities of each possible firing pattern in a set of 10 neurons are plotted against the observed probabilities; data pooled from all sets of neurons in culture 1; for reference, pictured in blue are the probabilities predicted by the independent model. (B) The distributions of predicted and observed sizes of spiking patterns.

For quantitative evaluations of model adequacy across multiple sets of neurons, multi-information ratio and Shannon-Jensen divergence were employed. Distribu-

tions of both these measures for the baseline recording from culture 1 are presented in Figure 3.5 (compare with Fig.2 a) and b) of Schneidman et al. (2006); bearing in mind that the absolute values of multi-information or Shannon-Jensen divergence are not necessarily equivalent due to differences in recording parameters and preparation properties). Multi-information ratio, as explained in Section 3.2.4, is a measure of the extent to which the maximum entropy model can explain the compound correlation within a group of neurons. Values near 0 indicate that the maximum entropy model performs similarly to the independent model and higher-order terms are necessary to account for the 'missing' multi-information; while values close to 1 (or, equivalently, a 100%) indicate that pairwise interactions are sufficient to capture the correlation structure in the data.

As evident from Figure 3.5A, the Ising models provided noticeable improvement over the independent model for all sets of neurons. Yielding MIR values ranging from 67% to 89%, with a mean of 79%, they captured a substantial portion of multi-information, although they did not reach the performance expected of a perfect model. However, in order to fully appreciate the meaning of obtained multi-information ratio, the length of recordings and the distributions of firing rates and correlations also need to be taken into account, as these importantly influence the extent of the limited sampling bias, which in turn lowers the attainable MIR value (or, in other words, results in underestimation of the model MIR). Therefore, in addition to the multi-information ratios for the modelled distributions, the ratios for re-fit re-sampled data are also reported here, to provide a reference of practically attainable MIR values.

Note that for this surrogate data-set the MIR values saturate with increasing multi-information; this is most likely due to the fact that the low multi-information regime corresponds also to lowest rates and / or correlations, where the uncertainty of entropy estimation is higher. A comparison of the model MIR with the re-fit re-sampled data-set indicates that the approximate 80% of explained multi-information could in principle be surpassed by a more accurate model, to the level of 96%. By rough calculation, that puts the performance of the Ising model in terms of capturing multi-information at $\approx$ 83%. Thus, although accounting for the majority of it, the pairwise maximum entropy model failed to capture all of the compound correlation structure of the data.

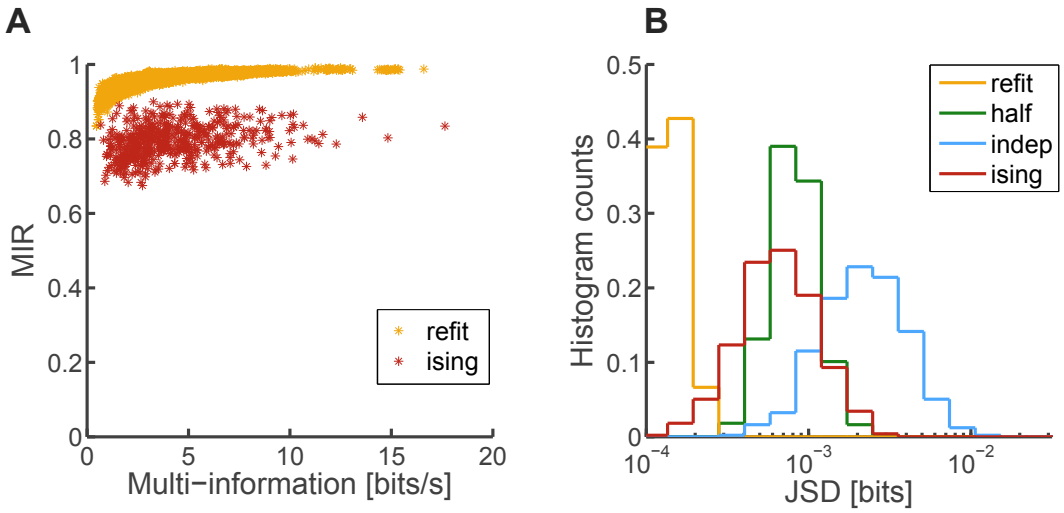

Figure 3.5: The distributions of the measures of model fit quality in the baseline recording in culture 1. (A) The model multi-information ratio plotted as a function of multi-information for each set of neurons; also plotted for reference the multi-information ratio of re-fit re-sampled dataset. (B) Normalized histogram counts of Jensen-Shannon divergences of model fits across all sets of neurons; also plotted for reference are histograms of Jensen-Shannon divergences of independent models for same groups of neurons, as well as divergences of distributions derived from halves of the data and those of the re-fit re-sampled dataset.

This result and the MIR values obtained match with those reported by Tang et al. (2008) for dissociated cortical cultures data binned at 20ms (approx. 81%, see their Fig. 4B), and surpasses their result for same data binned at 4ms (approx. 68%, see their Fig. 4D). Similar multi-information ratios were also reported by Yu et al. (2011) for Ising model fits to LFP avalanche patterns in *in vivo* data. On the other hand, the present results did not reach the $\approx$ 90% MIR obtained in the studies of retina (Schneidman et al., 2006), or the 92-93% exhibited by *in vivo* cortical spiking patterns in response to visual stimulation (Yu et al., 2008). Given the fact that Schneidman et al. (2006) report higher firing rates but comparable correlation coefficients between units to the present study, it is possible that in their data the higher-order correlations have less of an impact than here. It is debatable however to what extent these discrepancies in MIR are a reflection of the true impact of high-order correlations, and to what extent an exposure of the limitation of the methods and the measure itself. It bears reminding that the MIR gives a relative score with respect to the independent model, i.e. it reports what *fraction* of multi-information is explained by the pairwise maximum entropy model. An implicit assumption in here is that the multi-information is a robust measure of compound correlation; this assumption however hinges on the assertion of stationarity. Therefore, to better understand the problem and to examine

the true sources of model failure, it is informative to consider alternative points of view of model fit quality.

As explained in Section 3.2.4, as a complementary measure the Jensen-Shannon divergence for probability distributions was used. Similarly to the MIR, this measure also yields values between 0 and 1, with 0 indicating lack of similarity, and 1 indicating perfect agreement. The distributions of JSD values across all sets of neurons for both independent and maximum entropy models are presented in Figure 3.5B, along with the distributions of JSD values for the re-fit re-sampled data and the half-half comparison. The latter two in particular deserve attention, since they provide the reference points necessary for better understanding how accurate the model was. The re-fit re-sample dataset, as mentioned before, provides an estimate of how accurately the distributions can be recovered in the regime under which the data was collected. Here again it is evident that the Ising model underperformed in recovering the original data distributions, confirming the observations from the MIR analysis.

The half-half data-set had a similar purpose to the re-fit re-sampled one, with the important difference of also accounting for nonstationarities. Since spiking probability distributions were compared between the first and the second half of the recording, the Jensen-Shannon divergence reflected not only the issue of limited samples (which, in this case, was half of the original recording), but also the issue of possible slow fluctuations in firing patterns or influence of stochasticity. Interestingly, the distributions of JSD values of the pairwise model (Figure 3.5B, maroon) and the half-half data-set (Figure 3.5B, green) were very similar, and in fact in some sets the pairwise maximum entropy model provided a better match for the observed distributions than was the similarity between halves of collected data. This suggests that most of the failure of the model was not necessarily due to higher order correlations in the data, but should rather be attributed to the fact that a stationary model was used for data that was not fully stationary.

To further examine the sources of model failures, and specifically whether there was any relation between lack of stationarity and shortcomings in explaining the compound correlation, in Figure 3.6 the half-half JSD values are plotted against the model JSD, MIR and multi-information values. Panel B of the above-mentioned figure demonstrates that the MIR was not related to stationarity (Pearson correlation for the plotted values 0.11); however, as evident from panel C, the multi-information

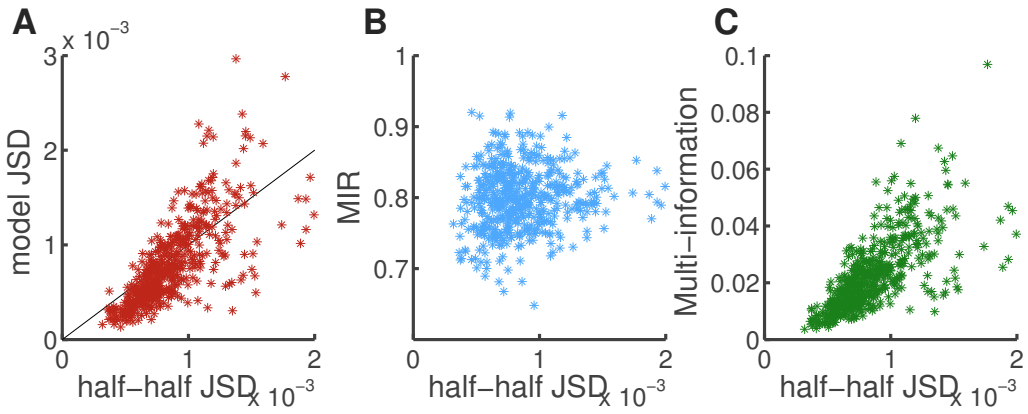

Figure 3.6: The relationship between recording stationarity and model fit quality measures. (A) The JSD of the pairwise maximum entropy model for each set plotted as a function of the JSD between distributions derived from halves of the recorded data. (B) The MIR plotted as a function of the JSD between distributions derived from halves of the recorded data. (C) The multi-information plotted as a function of the JSD between distributions derived from halves of the recorded data.

itself was. A comparison between panels A and C of Figure 3.6 shows that the extent of similarity of halves of the recording similarly predicted both the model JSD and the multi-information available (correlation coefficients, respectively, of 0.73 and 0.7). This result directly demonstrates that multi-information is not a robust measure of compound correlation. Thus, although the MIR itself appeared independent of non-stationarities, within the fraction of the multi-information that the model failed to account for were both high-order correlations and nonstationarities. Therefore, it is proposed here to focus on the JSD and its relation to the reference distributions rather than rely on the MIR scores.

It is unfortunately difficult to directly compare these results to other studies, as most of them employed only a subset of the measures / methods used here. For example, Tang et al. (2008) focus on multi-information ratio as a performance measure and do not report Jensen-Shannon divergences. Schneidman et al. (2006) on the other hand report JSD, but do not include any form of stationarity check. Interestingly, the Jensen-Shannon divergences distributions of the present work appear to match those reported by the latter study for the Ising, but not independent, model (compare Figure 3.5B with Figure 2b) of Schneidman et al. (2006)). This does not necessarily imply identical model performance, as the Jensen-Shannon divergence scores are also affected by limited sampling bias and the extent of the bias likely differs between the

two studies (as a reminder, Schneidman et al. (2006) use retinal data, which exhibits significantly higher firing rates of units).

Also Yu et al. (2011) reports the JSD; furthermore, the authors employ the same data-to-data comparison as here, and in addition a comparison with the Dichotomized Gaussian (DG in short, Macke et al. (2009)). In their analyses the pairwise maximum entropy model is shown to exhibit Jensen-Shannon divergences clearly larger than both the DG model and the half-half comparison (see Fig. 11 B and F of Yu et al. (2011)), although this difference is at least an order of magnitude smaller than that between the pairwise model and an independent one. However, the extent of improvement by the DG over the pairwise maximum entropy model varies depending on the data used (avalanche patterns versus instantaneous local field potentials versus spiking patterns) and the authors dissociate the high-order effects of temporal correlations (present in the avalanche patterns) from the high-order effects induced by thresholding. Here, in contrast, the Ising model exhibits performance comparable to the predictability of halves of the data but inferior to the predictability of re-fit re-sampled data. It appears therefore that the impact of higher-order correlations related to nonlinearities was smaller here than in the *in vivo* data-set of Yu et al. (2011) (a result consistent also with their analysis of relationship between high-order effects and rates and correlations). However, the discrepancy between the half-half data and the re-sampled one indicates temporal effects distinct to those uncovered by Yu et al. (2011) (or Tang et al. (2008)). The nonstationarity present in the 15-minute recordings and affecting the model fits here is not related to the fine-scale structure of avalanches or bursts but rather to the irregularity of bursting (see the example raster plots in Chapter 2, Figure 2.9 and 2.10, where it is apparent that in some recordings the halves contain a noticeably different number of bursts).

Overall, the performance of pairwise maximum entropy models reported here was within the range of other studies in the field, with certain understandable differences due to varying levels of stationarity and synchrony in different types of data. These observations regarding model fit quality held across cultures and recordings. As can be seen in Table 3.1, the MIR values averaged across all sets in a given recording were close to 80% in all cases, just as in the example from culture 1 presented earlier. The variability of multi-information ratio across sets was also comparable throughout, staying at the level of 4-5% (as measured by the standard deviation of MIR scores).

In general, all the recordings and cultures displayed appreciable homogeneity in this case, indicating that the applicability of the Ising model was not preparation-specific.

	0-2h	2-20h	20-44h	20-50h	50-68h	68-92h
culture 1: first	$79 \pm 4\%$	$80 \pm 4\%$		$80 \pm 4\%$	$82 \pm 5\%$	$80 \pm 5\%$
second	$80 \pm 4\%$	$81 \pm 4\%$		$82 \pm 5\%$	$80 \pm 5\%$	$80 \pm 5\%$
culture 2: first	$78 \pm 4\%$	$80 \pm 4\%$	$78 \pm 4\%$			
second	$80 \pm 4\%$	$79 \pm 4\%$	$79 \pm 4\%$			
culture 3: first	$77 \pm 4\%$	$80 \pm 4\%$	$76 \pm 4\%$			
second	$80 \pm 4\%$	$76 \pm 4\%$	$79 \pm 5\%$			

Table 3.1: The average ($\pm$ standard deviation) values of the multi-information ratio across all sets in a given recording, for all consecutive pairs of recordings.

With regards to the JSD, in all cultures and all recordings the pairwise maximum entropy model performed on a comparable level to how well halves of the data predicted each other. As shown in Figure 3.7, in the majority of cases the Jensen-Shannon divergences were similarly distributed for the model-to-data and half-half comparisons.

In culture 2, the half-half divergence was on average smaller than the model one, due to a very narrow range of the distributions; however the distributions still overlapped with those of the models. Interestingly, in culture 3, for some recording pairs, the model clearly, although not by much, surpassed the predictability of the data. These minor differences between the data-sets and the fact that there were also differences in the absolute levels of Jensen-Shannon divergence between cultures are resemblant of the variability across various studies and preparations discussed in earlier paragraphs. It appears thus that the cultures under study also exhibited a degree of heterogeneity in terms of stationarity and prevalence of higher-order correlations. Most likely, this variability is an effect of the heterogeneity of hippocampal cultures mentioned in Chapter 2, such as slightly different degrees of cell clustering on the array etc. However, the discrepancies between cultures were subtle and do not in themselves indicate any substantial issues with the choice of the model.

To conclude, the Ising models provided a satisfactory fit for the present neural data, matching the predictability of halves of recordings; this confirms and extends the observations from other studies (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008; Yu et al., 2008). It is worth remarking here that demonstrating adequate fit quality was important for further analyses, to ensure that the results did not simply

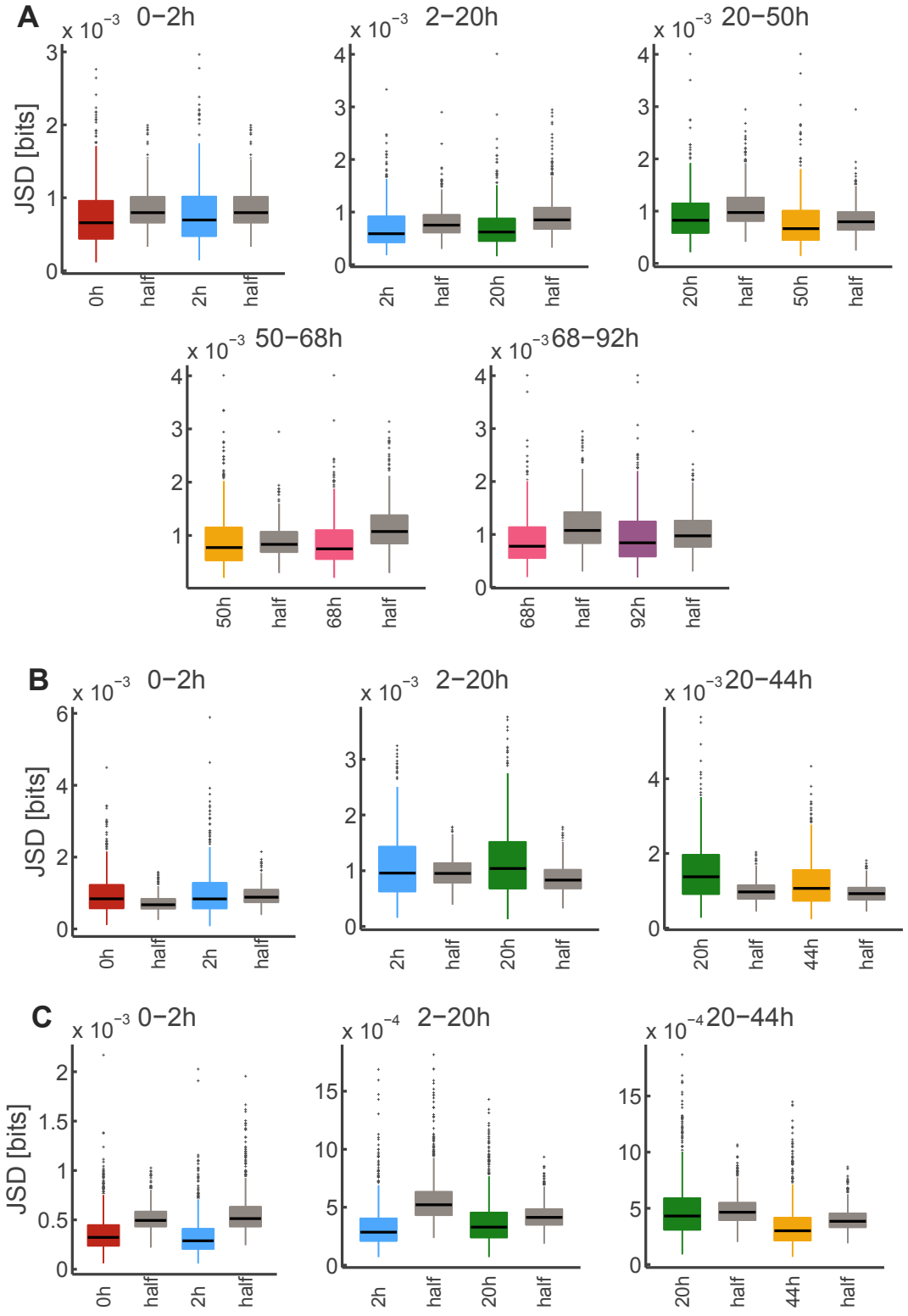

Figure 3.7: The distributions of Jensen-Shannon divergences of the model-to-data and data-to-data comparisons in each recording, for all consecutive pairs of recordings. (A) Boxplots of the JSD values for culture 1. (B-C) Same for cultures 2 and 3.

reflect the properties of Ising models, but that these properties meaningfully translate to the underlying ensembles of neurons (Yu et al., 2008).

3.3.2 Parameters of the fitted models: general properties and interpretation

Thus it transpires that the Ising model provided a reasonable description of the activity of groups of neurons. Importantly, this description, aside from being concise and analytically tractable, was not a trivial re-mapping from firing rates and correlations. In order to highlight this fact, it is useful to examine the parameters of the fitted models. To that end, Figure 3.8A and B shows scatter-plots of the fields and interactions against the corresponding firing rates and Pearson correlation coefficients of individual units and unit pairs. Several interesting features stand out in these images. First of all, the relationship between data marginals and model parameters is clearly not a linear one, an observation that conforms with previous reports (Schneidman et al., 2006; Tang et al., 2008). In particular, the distribution of interactions with increasing correlation appears to saturate (here at approximately 1), and in turn the spread of correlations tapers down at around 0 with decreasing interaction (compare Figure 3.8B with Fig.3 d) of Schneidman et al. (2006) and Fig.5 A of Tang et al. (2008), which show a qualitatively similar picture; also see Yu et al. (2008) for a different result, albeit in a very narrow regime of correlations and interaction values, which suggests the authors were dealing with a stimulus-driven regime rather than one dominated by inter-neuronal dynamics). The distribution of fields also exhibits signatures of this trend (Figure 3.8A), although it is confounded by other effects (see next paragraph). This characteristic nonlinear re-mapping reflects one of the key useful properties of the pairwise maximum entropy model: the fact that, contrary to its 'pairwise' name, it takes into account simultaneously the activities of all neurons within a group and assigns each unit a set of parameters that best describe its effect on the rest and its share in the collective behaviour. This property allows for factoring out common influences in neuronal activity and co-activity, such as for example common drive from a shared neighbour – something that correlations are notorious for ignoring (hence the saturation trend observed in Figure 3.8); in that sense, interaction parameters are a more meaningful measure of functional connectivity than correlations, which can often be spurious.

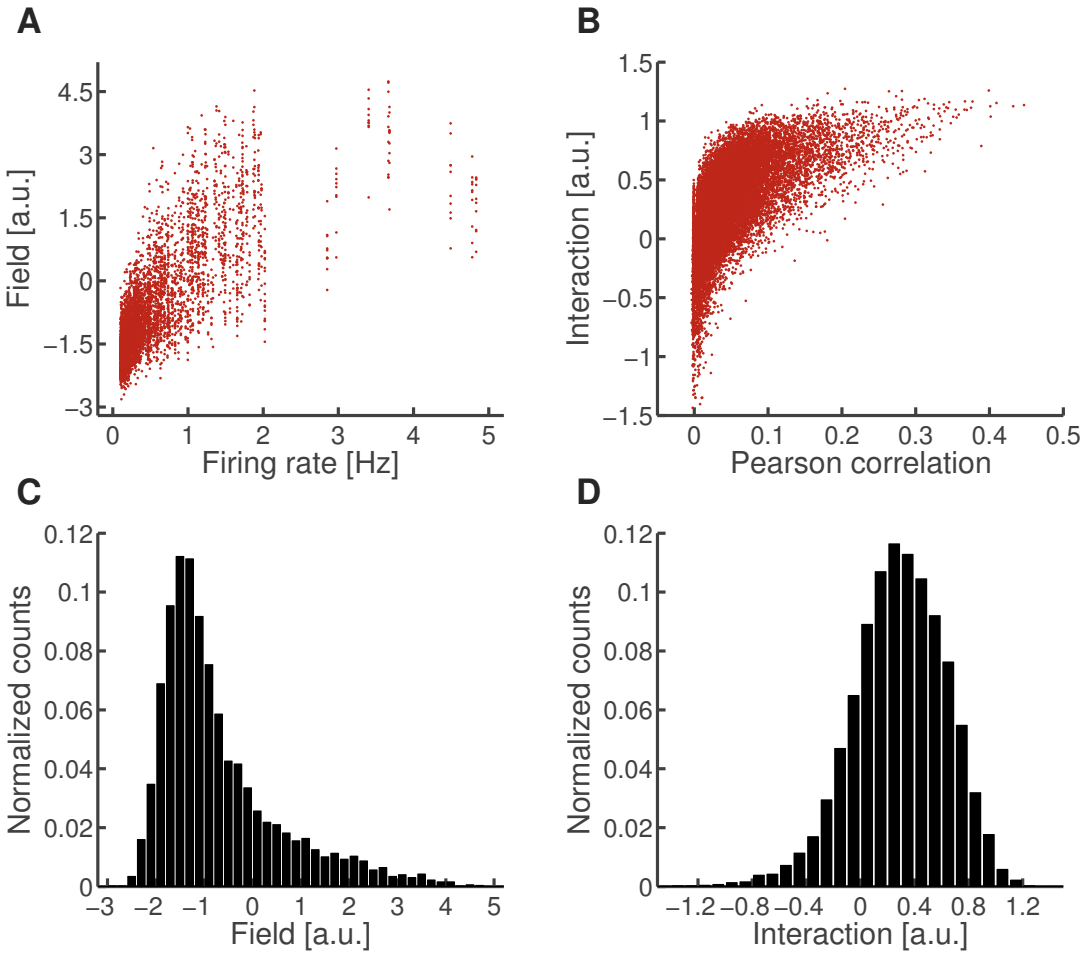

Figure 3.8: The distributions of the parameters of fitted models. (A) Scatterplot of model field parameters against the corresponding firing rates of neurons, pooled across all sets from culture 1, baseline recording. (B) Scatterplot of model interaction parameters against the corresponding Pearson correlation coefficients of pairs of neurons, pooled across all sets from culture 1, baseline recording. (C) Histogram counts of the values taken by the field parameters pictured in panel A. (D) Histogram counts of the values taken by the interaction parameters pictured in panel B.

The second observation to stand out from Figure 3.8 is also related to the fact that the Ising model simultaneously takes into account all neurons in a group. In panel A multiple vertical bands of points are apparent, clearly visible for the few most active neurons (with firing rates above 2Hz). As the reader might recall, the group sampling was done such that each neuron was chosen approximately 10 times; therefore for each unit there will be also 10 values of the fitted field parameter. The value of this parameter can be different in different groups, because, just like the interactions, fields are also functional parameters, inferred in relation to the set of neurons included in the calculations. Thus, for example, if in one of the groups a

strong interaction drives a chosen neuron, and in another group this interaction is not part of the ensemble, the field parameter would then need to be higher in order to explain the high activity of said neuron, effectively 'absorbing' the missing influence. However, there are no such clearly discernible vertical bands in Figure 3.8B. This might simply be due to the fact that there are fewer cases of the same pair of neurons being fitted multiple times, but a comparison with literature suggests that another factor might be that interactions are less affected by interpretational ambiguity (Yu et al., 2008). Since the latent units tend to be 'absorbed' into fields, for the interaction parameters the main reason for ambiguity would be a polysynaptic connection – however, in others' and in the present study the time bin was purposefully chosen so as to avoid including polysynaptic connections; still, a certain degree of variation between groups might still be expected.

Thus, although the functional form of the probability distribution suggests an elegant intuitive interpretation for the model parameters (i.e. fields quantifying the intrinsic excitability of each unit, and interactions quantifying the functional connectivity between pairs of units, similarly to the interpretation within the framework of the Hopfield model (Hopfield, 1982)), such clear one-to-one mapping could only be envisioned for ensembles where a complete set of interacting neurons is fitted with a model and there are no latent units which might need to be 'absorbed' into the fitted field parameters (for a brief discussion on including more units, see Appendix B. This limitation, in fact, applies to nearly any choice of measure, be it model parameters or directly measured excitatory potentials; namely, in order to infer accurate intrinsic properties of neurons from their activity, activities of all relevant units need to be known and included in the inference. However, given the complexity of neuronal activity and the density of neurons and their connections, such a feat is currently impossible.

Furthermore, such an approach misses two crucial issues: first, that just as all neurons need to be involved in inference of intrinsic properties, then all would then need to also be included in any modelling or attempts at functional predictions of group behaviour, which again would be prohibitively computationally complex; and second, the fact that functional groups of cells can overlap and share units. This is the advantage of employing models such as the Ising one; the parameters essentially describe the functional role of neurons within the group and how they and their in-

teractions contribute to creating collective patterns. The ambiguity brought with this approach is not necessarily a drawback, but rather an asset, allowing to appreciate the multiplicity of roles that neurons can take. And, paired with the approach of extensively sub-sampling the population, this can still uncover the deeper regularities of functional structure of large networks. To employ the example brought up above: observing that the individual units' field differs from group to group allows to correctly infer that there are missing interactions; now, as will become clear later in this work, it transpires also that units with high values of the field parameter also tend to have somewhat stronger, on average, connections within groups - taken together, these two facts allow to infer a certain inhomogeneity in the population, with some neurons appearing as 'hubs' of activity / connectivity (an observation that is also consistent with the results reported by Yu et al. (2008)). Consider also the fact that although a given neuron could potentially play vastly different roles in different sub-groups, the parameter distributions are not quite so arbitrary. That is, highly active neurons do tend to have high values of the field parameter and highly correlated ones of the interaction parameter. This rough conservation of rank hints at the possibility that neurons tend to play similar roles in different groups, and further corroborates with the notion that the high-firing units could be the leaders of local communities, perhaps forming a 'rich club' (Buzsáki and Mizuseki, 2014) within a small-world populations of neurons (Yu et al., 2008).

In this context, it is interesting to examine how the interactions might depend on the distances between the corresponding pairs of units. As shown in Figure 3.9, there was a small to moderate negative correlation between functional connectivity and distance on the array, depending on the preparation. Thus it appears that neurons closer by had higher interaction parameters and were more likely to fire together. It is tempting to speculate that this is yet another manifestation of the small-world nature of dissociated cultures. Such structure would be in line with what has been found in other studies (Schroeter et al., 2015). In particular, this could explain the nature of burst propagation in *in vitro* preparations, where activity spreads across the array in a wave-like fashion, activating groups of electrodes in a sequential manner (in-house observations, as well as other reports Marom and Shahaf (2002), Segev et al. (2004)). Since the Ising model is intended to capture only the stationary distributions of spiking patterns, no conclusions can be drawn here directly about the spatio-temporal

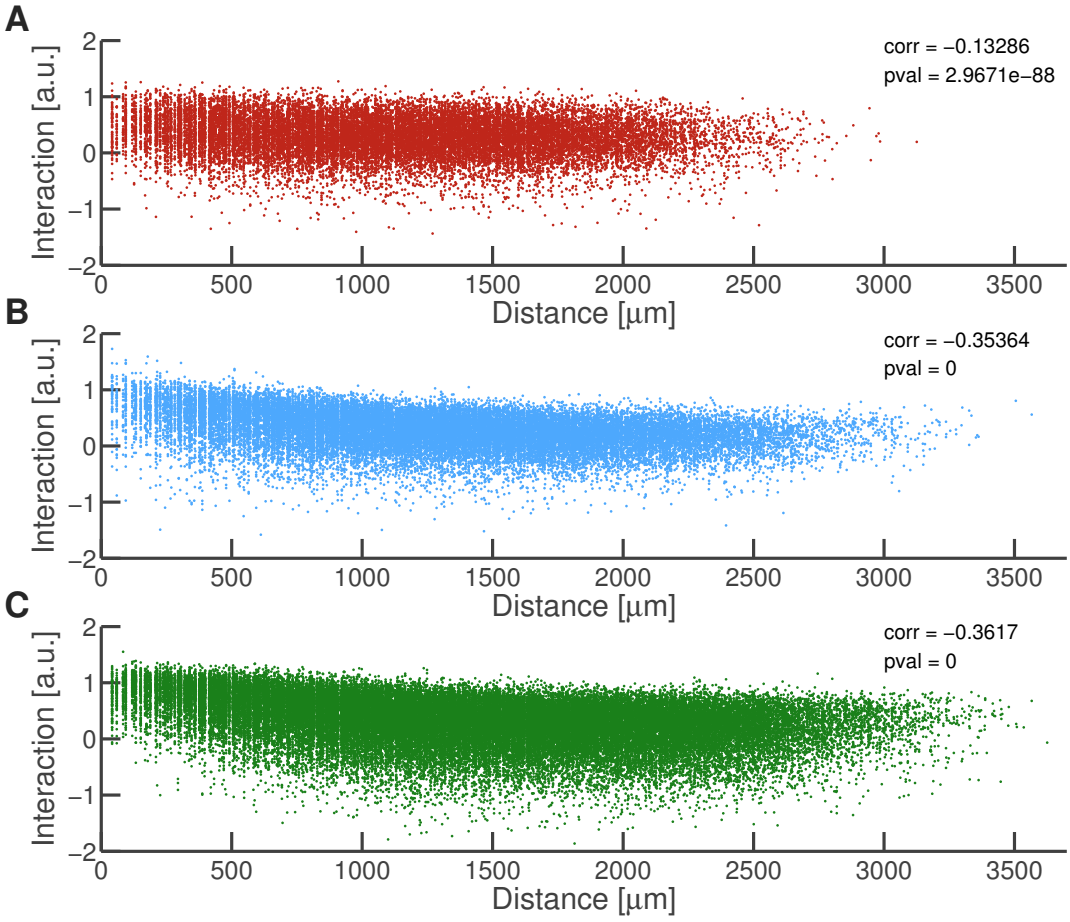

Figure 3.9: Distance dependence of interaction parameters. (A) A scatterplot of interaction parameters of each pair of units as a function of the distance between said units; culture 1, baseline recording. Inset the correlation coefficient between interactions and distance, along with the estimated p-value of the calculated correlation. (B-C) Same for cultures 2 and 3.

propagation of activity; nevertheless, the structure of locally clustered interactions is congruent with observations of the modular structure of bursts, which propagate in time preferentially through well-connected hub nodes (Schroeter et al., 2015).

A final observation from Figure 3.8 is that the distributions of both fields and interactions show deviations from the reported literature (Schneidman et al., 2006; Tang et al., 2008). First, although in the majority of cases the fields are negative, here in contrast to other studies a significant portion of values falls above zero (Figure 3.8C). Second, the interactions are predominantly positive (Figure 3.8D), unlike Tang et al. (2008) and similar to Schneidman et al. (2006), but the long tail of the distribution is in the negative values, in contrast to Schneidman et al. (2006). This is not surprising, as the type of preparation was different from Schneidman et al. (2006), and the high-

resolution recordings allowed for a more thorough and accurate picture of activity than in Tang et al. (2008). In particular, employing the 4096-channel MEA allowed to reduce the influence of a commonly observed bias in electro-physiological recordings, wherein the tendency is to pick up on the activity of strong and highly-active units, ignoring the much more common low-active neurons (Lütcke et al., 2013). The inclusion of less vocal units in present analysis is for example a likely contributor to the fact that positive values of the field parameter were obtained in some cases.

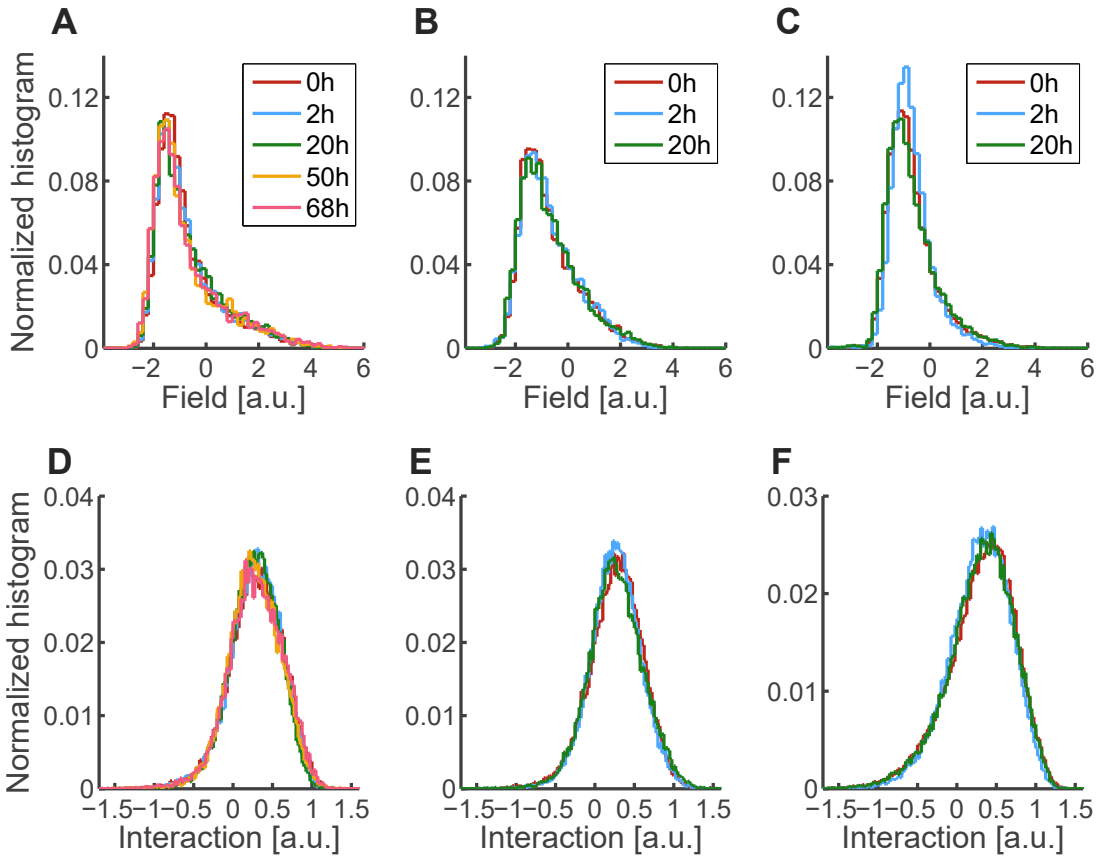

Figure 3.10: The distributions of the parameters of fitted models across recordings and cultures. (A) Normalized histograms of the values of field parameters pooled across all sets in a recording, for each baseline recording from a consecutive pair compared in culture 1 (see legend). (B-C) Same for cultures 2 and 3. (D-F) Same for interaction parameters (legends omitted for clarity, same as in panels A-C).

All of the above observations largely held for the remaining recordings and cultures. As summarized in Figure 3.10, the distributions of parameter values were qualitatively very similar across the data-set, although a closer examination reveals that the degree to which distributions were skewed differed between preparations. Notably, in the present study in all cases negative interactions and positive fields were

present, and distributions were relatively smooth (compare with Fig.3 b) of Schneidman et al. (2006) where a bimodal histogram of field values is presented). Interestingly, the pooled distributions of parameter values were very stable across recording points, mirroring the population stability discussed in Chapter 2 and illustrated in Figure 2.14).

3.3.3 Exploring comparisons of models across time

The pairwise maximum entropy models offered a particularly useful tool for the purpose of the study, in that they captured the essence of group behaviour and mapped it onto a set of individual and pairwise parameters in a manner that provided a mechanistic model of collective behaviour of neurons. This supplied the opportunity to probe the relationship between individual parameters, group behaviour and stability.

First, as a detailed example, in Figure 3.11 a single group of 10 neurons is pictured at two different time points, 0h and 2h. It is clear from both visual inspection and the r^2 value that the model parameters were largely conserved between recordings. In fact, this particular group of units was more stable parameter-wise than most other chosen ensembles, as indicated by the fact that its coefficient of determination was in the upper 95^{th} percentile (see also Figure 3.12A, indicated by the vertical red line). The spiking pattern distributions did not, however, reflect such similarity. The Jensen-Shannon divergence value was in the upper 97^{th} percentile, which means that most groups of neurons exhibited a better match (note that JSD close to 0 indicates similarity, and the higher the value the worse is the fit between distributions). Therefore, it appears that the two measures of model similarity report inconsistent results.

For a systematic examination, in Figure 3.12 the distributions of the values of r^2 and JSD and their relationship across all groups of neurons from culture 1 are pictured, for the recordings at 0h and 2h. Panels A and B allow a better visualisation of the example group from Figure 3.11 in the context of the rest of the data, highlighting the counterintuitive result of distribution dissimilarity despite parameter similarity. Examining directly the relationship between the r^2 and JSD values further confirms the above observations. As shown in Figure 3.12C, there was no appreciable correlation between those measures, and many groups of neurons exhibited similar behaviour

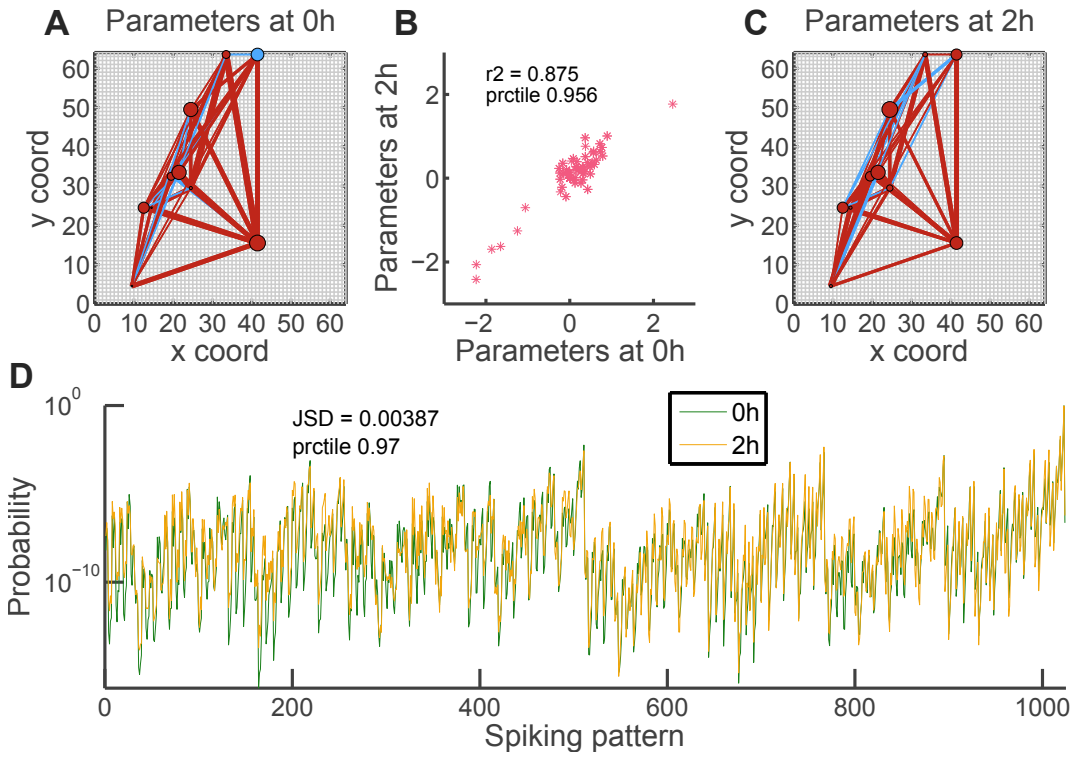

Figure 3.11: Example comparison of model parameters and pattern distributions between two time points, illustrating the nontrivial issue of the choice of appropriate similarity measure. (A) Parameters of the fitted model of a single group of 10 neurons at 0h, pictured over the MEA layout; line widths and circle diameters are proportional to corresponding parameter values; blue indicates a negative sign and maroon a positive one. (B) Scatter-plot of parameter values at the two time-points, 0 and 2h; the coefficient of determination for the parameters is reported, as well as the percentile of this r^2 value amongst all the groups of neurons. (C) Parameters of the fitted model for the same group of neurons as in panel A, but from a second recording at 2h; same notation applies. (D) The resulting pattern probability distributions for the chosen group of neurons at the two considered time-points; the Jensen-Shannon divergence is reported, as well as the percentile of this JSD value amongst all the groups of neurons (note that the bigger the value of JSD the smaller is the similarity between distributions).

to the one detailed in Figure 3.11. There was also no correlation when either or both sets of measures were log-transformed.

The lack of correlation between r^2 and JSD demonstrates that it was not predictable in a straightforward manner from the parameter changes alone whether groups were stable or not. This lack of predictability is due to the fact that even though the pairwise maximum entropy model provides a unique mapping between parameters and distributions, this mapping is nonlinear and complicated (hence the whole field of study of spin glasses). In fact, comparing parameters between time points is not a reliable measure of stability; nor is comparing the spiking pattern distributions via the Jensen-Shannon divergence, as explained in more detail in the next chapter and

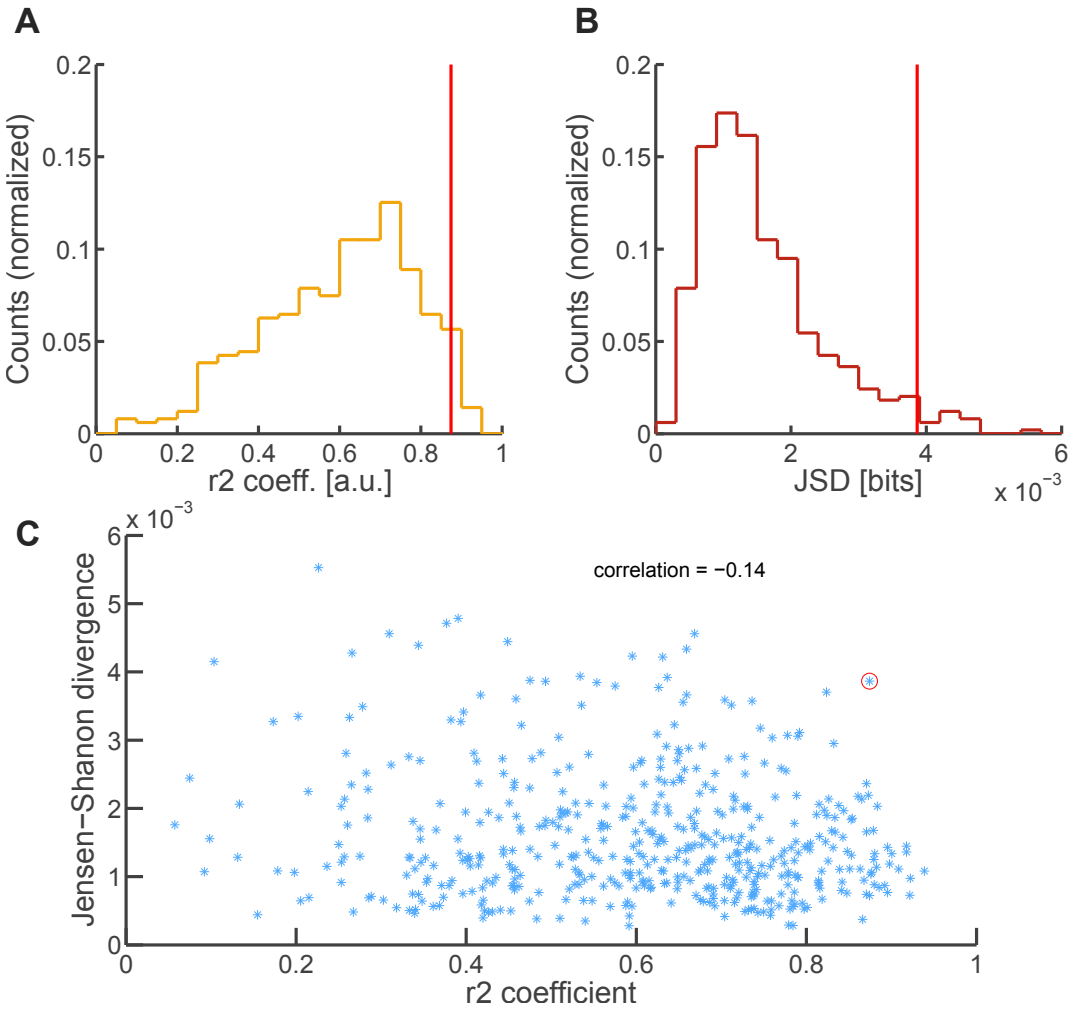

Figure 3.12: The distributions of model comparison measures: coefficients of determination of model parameters and Jensen-Shannon divergences of predicted distributions of spiking patterns (culture 1, comparison of recordings at 0h and 2h). (A) A histogram of the values of the r^2 across all groups of neurons; red vertical line denotes the coefficient of determination of the example group pictured in detail in the previous figure. (B) A histogram of the values of the JSD across all groups of neurons; again, red vertical line corresponds to aforementioned example. (C) The relationship between the coefficient of determination and the Jensen-Shannon divergence. A scatter-plot of r^2 values against the JSD values for each group of neurons; circled in red is the example group.

in Crooks (2007). In addition, the above result serves to illustrate how nontrivial it is to understand emergent phenomena in groups of neurons, even within a simplified framework of a pairwise maximum entropy model.

The results reported above held for all consecutive pairs of recordings and all cultures used in this study, as summarized in Figure 3.13. In fact, the heterogeneity between the preparations served to further bring out the counterintuitive relationship between parameter similarity and distribution similarity. For instance, neuronal

groups from consecutive recordings from culture 2 were on average more resemblant parameter-wise than either recording pair from cultures 1 or 3.

This is clear from visual inspection, as most of the distribution mass of the scatterplot in panel B falls above $r^2 = 0.7$, while in panel C it falls below this value. However, distribution-wise it was culture 3 that appeared to best retain between-recording similarity, as indicated by the fact that the Jensen-Shannon divergence tended to fall on average below 0.001. In culture 2, in contrast, only the first pair of recording exhibited JSD values of similar distribution.

Thus, to recapitulate, neither parameter similarity, nor the likeness of the distributions can unequivocally inform as to the true similarity of the underlying model (and, by extension, the collective behaviour of neurons). This is due to the fact that the bare values of parameters do not linearly translate onto parameter importance within a group. For example, if a few parameters are highly influential over collective group behaviour, all other parameters might change between time points, resulting in high value of r^2, while the ensemble as a whole remains stable. In a similar vein, it is not *a priori* clear which spiking patterns are the crucial components of group behaviour, and it does not necessarily follow from their absolute frequency. Therefore, yet another approach is called for in order to understand group stability and its relationship with individual parameters. Information theory furnishes the necessary tools for the task, as presented in the following chapter.

3.4 DISCUSSION

In this chapter of the thesis, the pairwise maximum entropy model is proposed as a tool for dissecting the complex relationship between the parameters and interactions of single neurons and the collective behaviour of neural ensembles. This model has been previously demonstrated to provide an excellent explanation for synchronous activity of cells in the retina (Schneidman et al., 2006; Shlens et al., 2006), and has been then extended to a variety of neural data (Tang et al., 2008; Yu et al., 2008) and employed to infer functional connectivity in the visual cortex of the cat (Yu et al., 2008). Here, the results of previous research are further extended, to now include the activity of neurons recorded with high-density multielectrode arrays (Berdondini et al., 2005), a novel type of data with different parameter requirements.

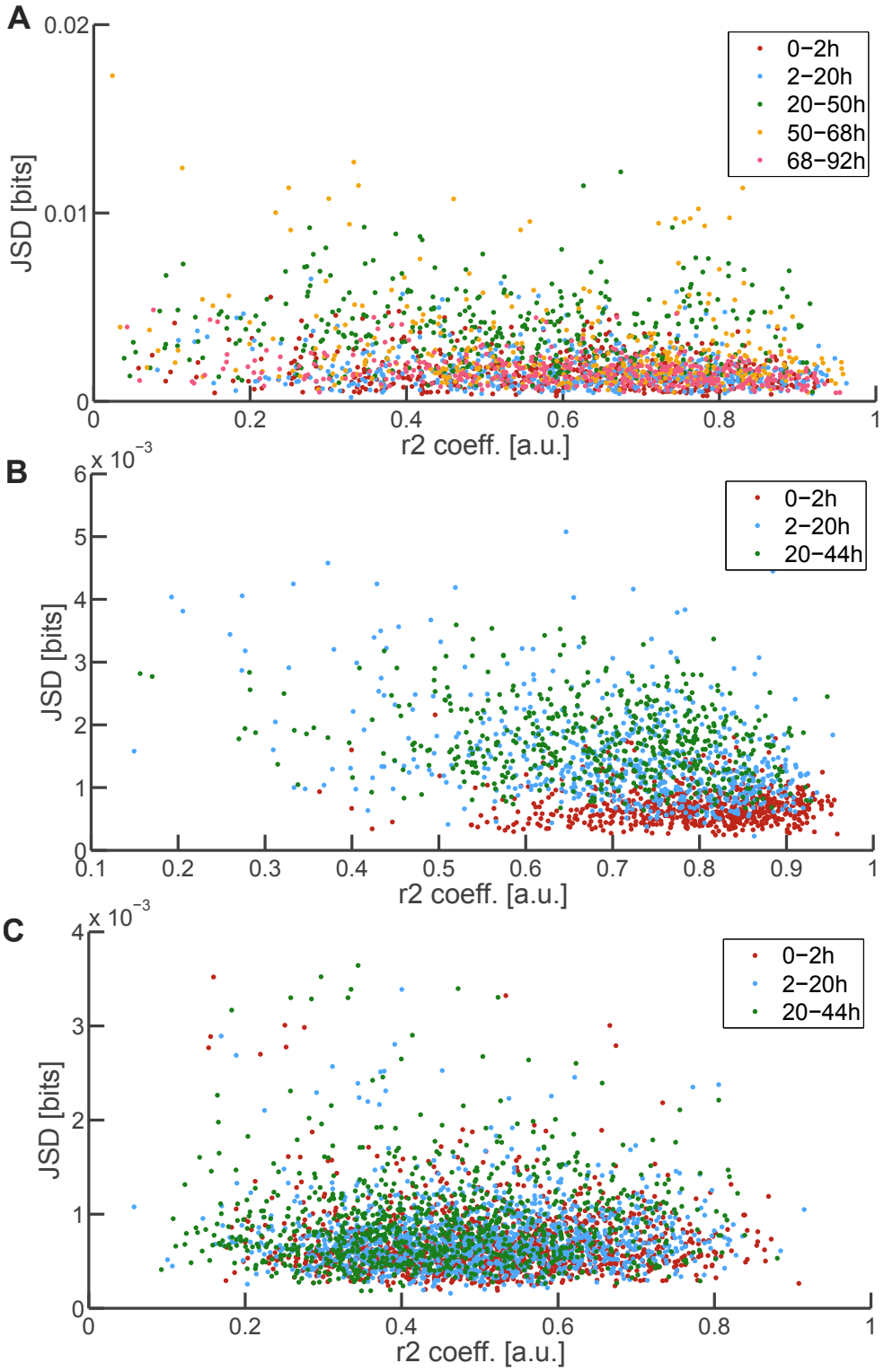

Figure 3.13: The relationship between the coefficient of determination and the Jensen-Shannon divergence for all consecutive pairs of recordings in all cultures. (A) A scatter-plot of r^2 values against the JSD values for each group of neurons from each consecutive pair of recordings in culture 1; color-code denoted in the legend. (B-C) Same for cultures 2 and 3.

Employing the multi-information ratio, the Jensen-Shannon divergence and additional comparisons to re-sampled data and data-to-data predictions of patterns, the model fits used in the present study are shown to fall broadly within the range of results reported by other researchers (Schneidman et al., 2006; Tang et al., 2008; Yu et al., 2011). Importantly, the pairwise maximum entropy models perform comparably to what can be achieved by comparing patterns derived from halves of the recordings, and an order of magnitude better than the independent model. Thus, they provide as close a match to the data as can be achieved without overfitting. Furthermore, introducing a comparison against a re-fit re-sampled dataset made it possible to reference the models against a perfect model in the same sampling regime. This has uncovered here a novel aspect of temporal dependencies in the data, since the half-half JSD distributions transpired to be significantly lower than that of the re-fit re-sampled JSD.

Since the previous research revealed limits in the predictive power of the pairwise maximum entropy model in certain types of data and applications (Tang et al., 2008; Ohiorhenuan et al., 2010; Yu et al., 2011), it is worth discussing the particular sources of the shortcomings of the model and their applicability here. Firstly, the model tends to fail when applied to temporally concatenated patterns (Tang et al., 2008; Yu et al., 2011). This is due to the fact that as time bins are concatenated, temporal correlations are re-mapped to spatial correlations, which often introduces beyond-pairwise terms. While models such as the Dichotomised Gaussian are able to accommodate those higher-order correlations (Yu et al., 2011), it was the view here that in such cases the method is reduced to being a low-dimensional data description, where parameters are likely confounded by mixing spatio-temporal effects, and thus not readily interpretable. Here, in contrast, the focus was solely on the statistics obtained from 5ms time bins. The rationale behind this was to dissociate the spatial domain from the temporal one, and to probe only the putative monosynaptic connections and their properties.

Secondly, the model has also been shown to fail due to higher-order correlations not related to temporal effects (Ohiorhenuan et al., 2010; Yu et al., 2011). These issues, however, appear to bear most relevance in *in vivo* settings, where the structure of neural circuits is much more complex (Ohiorhenuan et al., 2010), the sampling of units less exhaustive, and the operating regime of rates and correlations distinct from

dissociated cultures (Yu et al., 2011). Indeed, as discussed in detail in Section 3.3.1, the (non-temporal) higher-order interaction component was likely not dominant in the current setting, and in fact in many cases appeared negligible (see Figure 3.7).

The difference between the present results and *in vivo* report of Yu et al. (2011) presumably reflects the fact that in the cortex the local field potentials are more highly correlated across different sites than here, which the thresholding nonlinearity 'converts' to higher-order correlations. However, under their interpretation, the beyond-pairwise terms are not independent but exactly predictable from first- and second-order statistics. Therefore even if higher-order correlations were present in the data here, it should be inconsequential to ignore them, since they do not provide any extra information. In practice, however, caution should be taken when encountering model failures, as it is not entirely clear whether thresholding nonlinearity is the only source of beyond-pairwise terms; and under alternative explanations these terms can be important. An illustrative example can be found in Bettencourt et al. (2007) where authors consider a neuronal realization of an XOR gate, which exhibit zero pairwise but zero triplet mutual information and further demonstrate that such triplets can be encountered in neural data. Also Ohiorhenuan et al. (2010) report in their *in vivo* data significant higher-order correlations even in groups as small as 3 neurons. While these effects could be due to thresholding, alternative explanations cannot be excluded, for example a common unobserved input responsible for model failures (a likely scenario in the undersampled regime during evoked activity within a single microcolumn of the cortex (Ohiorhenuan et al., 2010)).

Although overall the higher-order interactions did not appear to contribute significantly to model failures here, a complication to the issue arises from the fact that the groups chosen for analyses were both relatively small (10, compared to the number of recorded neurons in order of several hundreds) and random. That is, it is not predictable from current results whether the observed lack of high-order effects should hold for larger groups, or specific non-random subgroups of units. As discussed in detail by Roudi et al. (2009b), it is impossible to conclusively extrapolate to larger populations if employed parameters lie in what is called the perturbative regime; and the group size, time bin and average firing rates of present work fall into that category. Since, however, neither a 'correct' group size can be *a priori* guessed, nor are all-population-spanning models feasible to fit, it was the approach here to

focus on the intermediate scale, under the assumption that in the self-organized systems that cultures provide, random sub-groups may constitute potential ready-to-use collectives.

Finally, as mentioned above, the use of re-fitted re-sampled dataset and half-half data comparisons reveals a temporal effect distinct from what was reported by other researchers (Tang et al., 2008; Yu et al., 2011). Although the Ising models perform here on the level comparable to predictability of data from its halves, in principle for the numbers of samples collected here it was possible to obtain even closer fits. This implies that there was some fluctuation in the statistics of observed patterns, possibly related to irregularity of bursting. It is unclear whether a similar effect was present in any of the studies referenced here, as only two of them employed data-to-data comparisons (Shlens et al., 2006; Yu et al., 2011), and none utilized re-sampling. Here the fluctuations are regarded as stochastic irregularities of bursting and averaged over the recording to obtain 15-minute snapshots of activity.

Treating the pairwise maximum entropy models as representations of groups of neurons is necessarily a simplification, one that ignores nonstationarity and higher-order interactions. It is nonetheless, in a sense, the first approximation to be used, since it is a parsimonious, tractable model, with pairwise structure mappable to functional connectivity, and conceptual roots in networks models of associative memory (Hopfield, 1982). Furthermore, the higher-order effects here are not dominant, and the models operate near the limit imposed by the predictability of the data. Importantly, the Ising model introduces a point of view superior to simple statistics of rates and correlations and offers the possibility of novel insights into synchronous behaviour (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008; Yu et al., 2008), which was the motivation behind employing it to study group stability. In particular, although the interpretation of parameters here is not as elegant and clear-cut as would follow from a full model – that is, the inferred parameters are a functional, rather than biological, measure of the strength of activity and co-activity, and somewhat ambiguous as to their locality – the re-mapping provided by the Ising model, paired with extensive sub-sampling of the population can still allow for valuable insights about functional structure of local and even global networks.

However, despite its relative simplicity and elegant mathematical formulation, the Ising model is not straightforward to draw conclusions from. Already at the relatively

small scale of 10 neurons the relationship between its parameters and the emergent collective behaviour is far from trivial. This becomes apparent when examining the similarity of groups across time (as detailed in Section 3.3.3), where the results of parameter comparisons and distribution comparisons appear to be at odds with each other (see Figure 3.13). Interestingly, although it is clear that some groups of neurons vastly change their parameters, their spiking patterns can appear almost unchanged; this further suggests that ensemble stability does not rely on individual stability (in line with the results of previous chapter). However, what is necessary to further understand how this comes about is a dedicated tool for dissecting model sensitivity to changes; this analysis is the core of the next chapter.

4

STABILITY, SLOPPINESS AND SPARSITY IN PARAMETRIC MODELS OF NEURAL ACTIVITY

4.1 BACKGROUND

Scientific and technological advances of recent years and decades have brought about a vastly changed landscape of research in biology. Development of such techniques as genome sequencing or 2-photon calcium imaging on one hand, and rapidly improving computational capacities on the other, resulted in the emergence of whole new areas of scientific inquiry, such as bio- and neuroinformatics and systems biology. A common denominator across many of these young disciplines is the increasing availability of vast amounts of novel data; and, informed by these increasingly detailed measurements, theoretical and computational models of growing complexity (Transtrum et al., 2015). Such abundance often-times brings with it even more questions than it answers; most notably, minutely accurate, biologically realistic models are notoriously difficult to interpret. As an answer to this challenge for human understanding, what has been recently proposed is the theory of sloppy systems (Machta et al., 2013; Transtrum et al., 2015): an information theoretic framework for analysing complicated predictive models, based on the Fisher Information Matrix. In essence, this approach allows for effective dimensionality reduction (Daniels et al., 2008; Machta et al., 2013; Transtrum et al., 2015) through a principled analysis of models' parameter sensitivity. The method in its original, more circumscribed form, was first introduced in biochemistry as a technique for addressing the problem of parameter indeterminacy in complicated, under-constrained models (Brown and Sethna, 2003). However, it has since helped uncover that surprisingly many complex scientific models exhibit a property named *sloppiness* (Daniels et al., 2008; Machta et al., 2013; Transtrum et al., 2015); that is, their behaviour is controlled by a combination of relatively few parameters. In fact, it has been suggested that the very fact that much of the physical world is amenable to concise and understandable scientific descrip-

tion is due to the fact that many of the real-world systems are sloppy (Transtrum et al., 2015). What is more, that same property has also been proposed to enable evolutionary stability and environmental robustness to changes, without concurrent loss of evolvability (Daniels et al., 2008; Transtrum et al., 2015).

Initially, the framework was introduced in order to aid parameter estimation and meaningful predictions from under-constrained, complex models of regulatory biochemical networks in eukaryotic and prokaryotic cells (Brown and Sethna, 2003). Such models typically encompass tens of ordinary differential equations, describing concentrations and reaction rates of interacting proteins, ribonucleic acids and other molecules. With tens of parameters to estimate, these models are however often informed by few and uncertain measurements, not allowing a robust fit, and leaving researchers to various *ad hoc* solutions, from improvised sensitivity analysis to educated guesswork (Brown and Sethna, 2003). Brown and Sethna (2003) propose a principled approach to alleviate this issue: starting from the relatively minimal assumption that the measurement errors are random Gaussian noise (and, implicitly, that the data comes from the proposed model), it follows that that the cost function of the model fit (i.e. the log-likelihood of data given the model) takes on the form of χ^2 (note that for Gaussian errors the maximum likelihood estimation of parameters is equivalent to the minimum least squares method); and this cost function can be then analysed by computing its Hessian matrix with respect to the model parameters at the point of the best fit solution. The Hessian thus obtained quantifies the sensitivity of the model fit to parametric changes, and can be further factorized to test for any regularities in its structure that could be exploited. Interestingly, the authors then demonstrate on an example biochemical regulatory network model the existence of such regularities. Namely, it transpires that the factorization yields an eigendecomposition of steadily decreasing eigenvalues, roughly uniformly in log-space, which means that there is a gradual separation of sensitivities in model parameters. Brown and Sethna (2003) term the high-eigenvalue combinations of parameters *stiff*, the low value ones *soft* (or *sloppy* in following work), and the general class of models exhibiting those properties *sloppy*.

Such particular structure of separated sensitivities has useful implications for modelling and understanding biochemical networks, or indeed any complex network with large numbers of parameters and data of limited availability (Brown and Sethna,

2003). The fact that only the first few factors of the Hessian are sensitive to changes means that the model has low effective dimensionality, i.e. its general behaviour will be the same under a wide range of sloppy parameters (Brown and Sethna, 2003; Daniels et al., 2008; Machta et al., 2013; Transtrum et al., 2015). Thus, being underconstrained by the data does not necessarily limit the predictive or explanatory power of the model (Transtrum et al., 2015), it merely means that only certain parameters or their combinations can or in fact need to be precisely determined; this can be explicitly accomplished with a novel method of dimensionality reduction (Transtrum et al., 2015), based on the discussed framework. However, this approach will only be successful in the cases where the underlying system itself is sloppy.

Remarkably, this has been shown to be the case in a vast array of systems. Since the original application described by Brown and Sethna (2003), a further 17 systems biology models have been demonstrated to posses the characteristic sloppy eigenvalue structure (Gutenkunst et al., 2007), along with several examples from other disciplines, including an artificial neural network (Transtrum et al., 2015). The framework has also been generalized to encompass all likelihood-based models (Transtrum et al., 2015), where the cost function needs not be postulated and the Fisher Information Matrix is the natural metric in model parameter space [1]. Furthermore, in their recent theoretical study, Machta et al. (2013) demonstrate that certain effective theories in physics can also be traced back to sloppy parameter space compression, as the scale of observation is coarsened. The above result along with the apparent ubiquity of both sloppy models and other dimensionally reducible models across science (such as mean field theory or renormalization group in physics) has brought researchers to ponder whether sloppiness is a fundamental property of most complex systems in the physical world (Transtrum et al., 2015); or rather, whether the hierarchical nature of physical theories in science is merely a reflection of the way the human mind works: we perceive and understand only those natural processes that we evolved to perceive. The answer to such profound epistemological questions is still beyond human reasoning; however, authors suggest that non-sloppy systems would be challenged to survive under Darwinian evolution (Transtrum et al., 2015).

1 The Fisher Information Matrix is defined as the expected value of the Hessian of the model log-likelihood, see Section 4.2.1. For cases such as in Brown and Sethna (2003), where no likelihood is defined within the model and is additionally imposed to be Gaussian, or for cases where likelihood is known to be Gaussian, the FIM at the point of best fit simplifies to Hessian of the χ^2 cost function at the point of best fit (Transtrum et al., 2015).

The link between sloppiness and evolvability is of particular interest in the present work, because of a parallel problem to the one considered here: how to reconcile global stability (such as mutational robustness of an organisms) with ongoing changes necessary for adaptability (such as genetic alterations). As reviewed by Daniels et al. (2008), sloppiness provides a natural and intuitive explanation for such robustness. The large space of insensitive parameters allows for a wide range of random mutations to take place, without impacting the phenotype and survival of most individuals. At the same time, across a population novel phenotypes can arise when random mutations occur in stiff parameters but at different areas of the vast sloppy space. As an example, Daniels et al. (2008) cite multiparameter models of the gene network responsible for segment polarity in the *Drosophila*. It transpires that under a surprisingly large array of randomly chosen parameters these models of gene expression produce a wild type patterning. And indeed, eigendecomposition conducted at one of the acceptable parameter sets produces a sloppy spectrum of eigenvalues (Daniels et al., 2008). Interestingly, the authors observe that also environmental robustness can be attributed to the separation of sensitivities, as exemplified by analysis of a model of temperature compensation in cyanobacteria. Altogether, evidence suggests that in many systems described by multiple micro-scale interactions sloppiness could be a key feature enabling stability without the need for a homeostatic control of every single parameter.

The above findings and observations have prompted a hypothesis for the present work: that also in models of neural activity sloppiness could provide a natural explanation for stability of emergent collective behaviour despite ongoing fluctuations in the parameters of single units. Conveniently, the Fisher Information Matrix is the natural metric tensor for energy-based models, such as the pairwise maximum entropy model employed here (Crooks, 2007; Amari and Nagaoka, 2007). In the present context therefore, examination of the matrix also emerges independently as the logical step of analysis of the relationship between model parameters and collective behaviour (see also Section 4.2 below). However, sloppy model research provides a useful conceptual framework for intuitive interpretation of results. From this point of view, the maximum-entropy model defines the likelihood of obtaining the data under the given parameters, for which the Fisher Information Matrix can be calculated and analysed (Transtrum et al., 2015). The matrix can be analytically derived in closed

form for the model class used here and calculated for each of the model fits directly (Crooks, 2007).

In the present chapter, examination of the Fisher Information Matrix of the pairwise maximum entropy models was used to test whether these models exhibited sloppy properties.

4.2 METHODS

The analyses described here were all implemented in Matlab by the author.

4.2.1 Comparing probabilistic models: theoretical foundations

As explained in detail in the previous chapter, the pairwise maximum entropy model is a type of statistical inference, or in other words it belongs to the broader class of probabilistic models. A probabilistic model is a mathematical description of a stochastic system, which provides a functional parametric form of the probability distribution of observing distinct states of the said system (Amari and Nagaoka, 2007)

$$p(x|\Lambda), \tag{4.1}$$

where Λ denotes a set of parameters. Probabilistic models can be inferred from minimal assumptions, as here, or postulated from hypotheses, such as e.g. when fitting an exponential function to the distribution of radioactive decay; regardless, however, of their genesis, they are all described by the branch of mathematics called information geometry (Amari and Nagaoka, 2007). This relatively young field of science deals with the mathematical properties of sets of probability distributions and thus provides a unified framework for analysing and understanding probabilistic models (Amari and Nagaoka, 2007).

What is of particular import here, parametric probability distributions form Riemannian manifolds, and as such they are naturally equipped with a coordinate system and a distance metric. The coordinate system is provided by the parameters of the model, and the tensor defining the metric is given by the Fisher Information Ma-

trix (Amari and Nagaoka, 2007; Machta et al., 2013) (also referred to from now on as FIM)

$$f_{lm}(\Lambda) \quad = \quad -\sum_{x} p(x|\Lambda) \frac{\partial^2 \ln p(x|\Lambda)}{\partial \lambda_l \partial \lambda_m}. \tag{4.2}$$

From the above expression the FIM can be immediately recognized as the expected value of the curvature of the model log-likelihood. Thus the Fisher metric tensor can be intuitively understood as the local estimate of the shape of the model manifold, as it provides a measure of change of the model in response to infinitesimal changes to each of its parameters (see Figure 4.1 for an illustrative example).

Having defined a metric tensor, the length of any curve in model state space can now be calculated as (Crooks, 2007)

$$L \quad = \quad \int_{t=0}^{\tau} \sqrt{\frac{d\lambda_l}{dt} f_{lm} \frac{d\lambda_m}{dt}} dt. \tag{4.3}$$

That allows for a principled measure of distance between model probability distributions as the shortest achievable path length (these are called geodesics and are the equivalent of straight lines in Euclidean space). The above is not an abstract construct applicable solely to mathematical objects; in fact, a whole class of physical systems in the field of thermodynamics described by the Boltzmann distribution behave according to the rules of information geometry (Crooks, 2007), and this includes the Ising model and spin glasses (see Crooks (2007) for a more in-depth treatment of the deep links between thermodynamics and information theory).

Let us recall here that Chapter 3 left unresolved the question of how to compare models across time, and it was pointed out that direct comparison of parameters was not truly informative of model similarity. It becomes clear now why this was the case – the models reside on a manifold and this manifold need not be uniform with respect to parameter coordinates. As illustrated in Figure 4.1, in certain regions of parameter space (the horizontal flat area, coloured dark blue) changing two different parameters by the same amount has a different impact on the predicted probabilities – in this case, $p(x = s_2|p1, p2)$ is quite sensitive to only one of the parameters, $p1$, but not the other. However, $p(x = s_3|p1, p2)$ is practically unaffected by either $p1$ or $p2$. In contrast, in the topmost green-coloured part it appears that here both parameters have similar impact on the $p(x = s_2|p1, p2)$, but now the $p2$ is the one that has more influence over

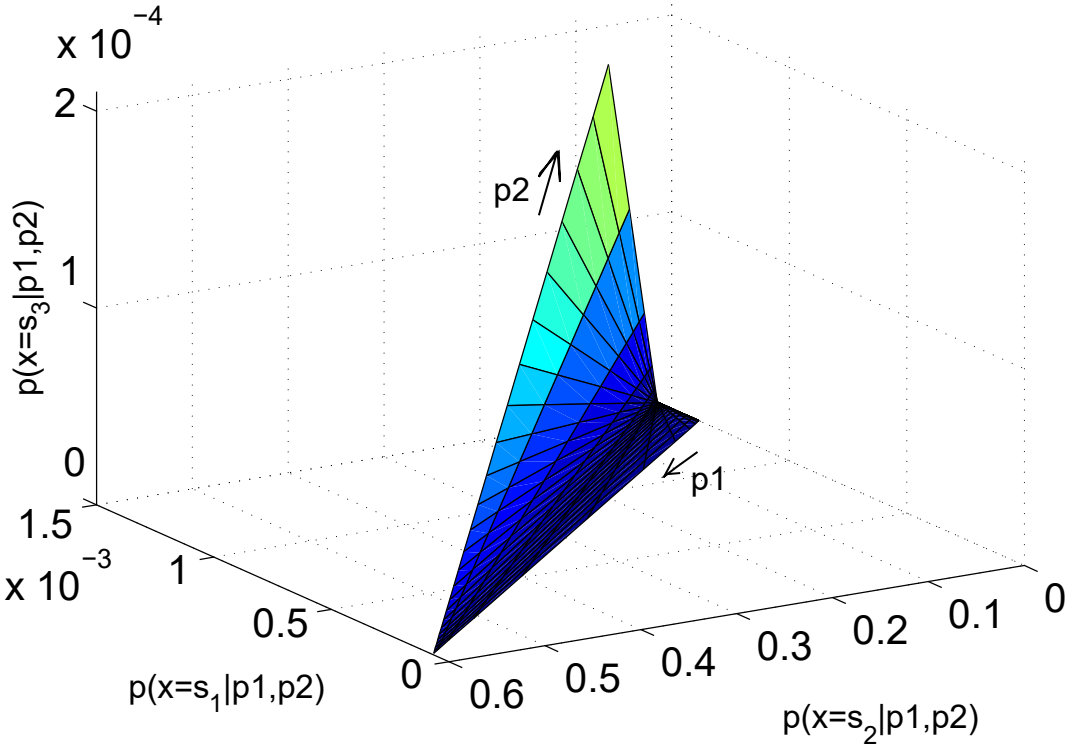

Figure 4.1: Illustration of the concept of a probabilistic model residing on a Riemannian manifold; for this purpose a two-dimensional cross-section of the manifold is presented, evaluated in 3D data space (the full manifold is 55D in 1024D data space). All but two parameters are kept fixed at the best solution values, and two interaction parameters are varied across the same range each. The vertices of the grid (x,y,z) give the probabilities of three distinct patterns s_1, s_2, s_3 as predicted by the model for each parameter combination; the directions of changes along the grid of $p1$ and $p2$ are shown in the plot. As can be seen from the irregular shape of the grid in the picture, those two parameters have a different effect on the measured probabilities in different regions of parameter space (note different scales of the axes).

one of the patterns, $p(x = s_3|p1, p2)$ (note, however, that changes along two of the axes are vastly smaller than the range exhibited by $p(x = s_2|p1, p2)$, so the situation is not symmetric between $p1$ and $p2$). Thus, already in this illustrative example there are hints of unequal sensitivity of the model to its parameters. Furthermore, this unequal sensitivity is locally specific, which provides an intuition as to why in some cases models with quite different sets of individual parameters can still be similar, while in other cases even a small change of parameters results in a vastly different emergent behaviour in a group of neurons.

Similarly, it was pointed out previously that also direct comparison of probability distributions via the Shannon-Jensen divergence is not a reliable measure of model similarity. Although the square root of the Shannon-Jensen divergence is considered a metric, it is not a Riemannian metric, and thus it cannot serve to measure path

lengths (Crooks, 2007). This can be intuitively understood by again considering the Figure 4.1 and focusing on the 'bent' shape of the manifold cross-section, which goes from horizontally flat to a steep rise along the z axis. What the JSD provides is the equivalent of measuring the distances between points on the surface solely by their distance along the axes. However, the true distance is one along the geodesic, because to get from one point on the manifold to the other, the modelled system needs to visit all the intermediate states that lie along the surface.

Thus, information geometry provides the theory and mathematical tools to meaningfully describe and analyse probabilistic models. It is, however, not trivial to calculate quantities such as path lengths in practice, since it is not known *a priori* what are the geodesics connecting each pair of models. And although Crooks (2007) propose an approximate computational solution, it is highly demanding in terms of processing, especially for the manifold dimensions considered here, and thus practically not viable.

However, since the focus of interest in the present work is to gain understanding of the role that the individual parameters play in group stability, it is not necessary to calculate path lengths; what suffices for the current purpose is analysis of the Fisher Information Matrix (Machta et al., 2013). The FIM provides a measure of the curvature of the log-likelihood landscape in the coordinate system of model parameters and as such it essentially quantifies the sensitivity of the model to parameter changes. Thus, analysis of the structure of the tensor matrix directly informs on the relationship between local and global properties of the model, allowing to side-step the issue of comparing models across time in order to infer this relation. In fact, knowledge of the sensitivity of the model with respect to individual directions in parameter space should be used to establish group stability, rather than parameter similarities or the Jensen-Shannon divergences.

4.2.2 The Fisher Information Matrix: calculation for the Ising model

For the pairwise maximum entropy model, the Fisher Information Matrix is easily calculated in closed form, as outlined below. The original Equation 4.2 can be rewritten as

$$f_{lm}(\Lambda) = \sum_x p(x|\Lambda) \frac{\partial \ln p(x|\Lambda)}{\partial \lambda_l} \frac{\partial \ln p(x|\Lambda)}{\partial \lambda_m} = \dots$$

$$\dots = \sum_x \frac{1}{p(x|\Lambda)} \frac{\partial p(x|\Lambda)}{\partial \lambda_l} \frac{\partial p(x|\Lambda)}{\partial \lambda_m}. \tag{4.4}$$

That leaves only the partial derivative with respect to any of the parameters to be calculated. Let us recall that the model is defined through the following equation

$$P(X_s) = \frac{e^{-\sum_i \lambda_i x_i - \frac{1}{2}\sum_{k\neq j} \lambda_{jk} x_j x_k}}{\sum_{\{X_s\}} e^{-\sum_i \lambda_i x_i - \frac{1}{2}\sum_{k\neq j} \lambda_{jk} x_j x_k}} = \frac{e^{-E_s}}{\sum_{\{X_s\}} e^{-E_s}}. \tag{4.5}$$

Although the parameters fall into two categories, the fields λ_i and the interactions λ_{jk} (associated with first and second order marginals respectively), the calculations are almost identical for either parameter type; for the sake of simplicity let us focus on λ_i. The derivation of Equation 4.5 with respect to one of the field parameters is easily conducted

$$\frac{\partial P(X_s)}{\partial \lambda_i} = \frac{\sum_{\{X_s\}} e^{-E_s} \cdot \partial_{\lambda_i}\left(e^{-E_s}\right) - e^{-E_s} \cdot \partial_{\lambda_i}\left(\sum_{\{X_s\}} e^{-E_s}\right)}{\left(\sum_{\{X_s\}} e^{-E_s}\right)^2} = \dots$$

$$\dots = \frac{\sum_{\{X_s\}} e^{-E_s} \cdot \left(-x_i \cdot e^{-E_s}\right) - e^{-E_s} \cdot \left(\sum_{\{X_s\}} \left(-x_i \cdot e^{-E_s}\right)\right)}{\left(\sum_{\{X_s\}} e^{-E_s}\right)^2} = \dots$$

$$\dots = \frac{e^{-E_s}}{\sum_{\{X_s\}} e^{-E_s}} \left[-x_i - \frac{\sum_{\{X_s\}} \left(-x_i \cdot e^{-E_s}\right)}{\sum_{\{X_s\}} e^{-E_s}}\right] = P(X_s)\left(-x_i + \langle x_i \rangle_{mod}\right). \tag{4.6}$$

In analogous manner it can be shown that for λ_{jk} the derivative yields

$$\frac{\partial P(X_s)}{\partial \lambda_{jk}} = P(X_s)\left(-x_j x_k + \langle x_j x_k \rangle_{mod}\right). \tag{4.7}$$

Replacing the derivatives in Equation 4.4 with the appropriate expressions leads to closed form for each entry in the Fisher Information Matrix. Let us consider as an example the entry for a pair of parameters $\lambda_l = \lambda_i$ and $\lambda_m = \lambda_{jk}$. This will take on the form

$$f_{i,jk} = \sum_{\{X_s\}} \frac{1}{P(X_s)} P(X_s)\left(-x_i + \langle x_i \rangle_{mod}\right) P(X_s)\left(-x_j x_k + \langle x_j x_k \rangle_{mod}\right) = ...$$

$$... = \sum_{\{X_s\}} P(X_s)\left(-x_i + \langle x_i \rangle_{mod}\right)\left(-x_j x_k + \langle x_j x_k \rangle_{mod}\right) = ...$$

$$... = \langle x_i x_j x_k \rangle_{mod} - \langle x_i \rangle_{mod} \langle x_j x_k \rangle_{mod}. \tag{4.8}$$

The above expression is resemblant of an element in a covariance matrix. Indeed, the Fisher Information Matrix can be understood as form of covariance matrix of the random variables associated with all parameters in turn (Crooks, 2007)

$$F(\Lambda) = \langle XX \rangle_{mod} - \langle X \rangle_{mod} \langle X \rangle_{mod}, \tag{4.9}$$

where $X = \{x_i..., x_{jk}...\}$ for the parameters $\Lambda = \{\lambda_i..., \lambda_{jk}...\}$; this expression for the FIM holds in fact for any probability distribution belonging to the exponential family (of which the Ising model is a member).

Since for the 10-unit groups the pairwise maximum entropy model is specified by 55 parameters (10 field parameters, sorted in ascending order, and 45 interaction

parameters, sorted as follows: $\{\lambda_{1,2}, \lambda_{1,3}, .., \lambda_{1,10}, \lambda_{2,3}, \lambda_{2,4}, .., \lambda_{2,10}, ..., \lambda_{9,10}\}$), the FIM here will thus be a symmetric 55x55 matrix of block form

$$F = \begin{pmatrix} f_{1,1} & \cdots & f_{1,10} & f_{1,\{1,2\}} & \cdots & f_{1,\{9,10\}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{10,1} & \cdots & f_{10,10} & f_{10,\{1,2\}} & \cdots & f_{10,\{9,10\}} \\ f_{\{1,2\},1} & \cdots & f_{\{1,2\},10} & f_{\{1,2\},\{1,2\}} & \cdots & f_{\{1,2\},\{9,10\}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{\{9,10\},1} & \cdots & f_{\{9,10\},10} & f_{\{9,10\},\{1,2\}} & \cdots & f_{\{9,10\},\{9,10\}}. \end{pmatrix} \tag{4.10}$$

where all entries can be directly computed from the model.

As a useful side note let it be pointed out here that technically, since these entries take on the form of covariances and $P_{mod}(X_s)$ is taken to be the underlying model of $P_{obs}(X_s)$, one could also compute the FIM directly from the observed data, as the expected values $\langle ... \rangle_{mod}$ and $\langle ... \rangle_{obs}$ should be equivalent. Could one then do away with the cumbersome procedure of model fitting and simply derive FIMs from the data? There are several counter-arguments to that notion, although the data-derived FIM can have its uses, particularly for long recordings and large number of neurons, where exact model fitting is computationally infeasible and one wants to test for sloppiness with increasing group size.

First, the data marginals and model marginals are truly equivalent only in the limit of infinite samples; however, for finite data both the observed probability distribution and the data marginals are contaminated by random noise. Since the model fitting procedure relies on the observed first and second order marginals, it is of course also affected by the noise; but the first and second order marginals are the ones least sensitive to noise, so the model depends solely on the most reliable data measurements. Therefore, it is expected that third and fourth order marginals calculated from the model would be a less noisy version of those calculated from limited observed data.

A second issue with disregarding model fitting is that there would be no verification of the quality of the model fit and appropriateness of the model for the data. The success of the Ising model relies on the fact that, as mentioned earlier, the modelled and observed distributions were compared and the model was able to account for the majority of multi-neuron spiking patterns (Schneidman et al., 2006). Forgoing

this step in application to novel data would leave researchers blind to the possibility that a different model might be required in such new context, with a different form of Fisher Information Matrix.

Last but not least, the fitted models supply a tractable description of group behaviour and parameters of functional properties of neurons, which are the focus of interest here. Albeit ambiguous to an extent (as discussed in Section 3.3.2), these parameters can be further analysed along with the FIM, as described in the following sections. In particular, parametric changes between time points can be related to FIM-inferred parametric sensitivity, something that could not be investigated performing the model fitting.

4.2.3 The Fisher Information Matrix: factorization

Once the FIM is obtained for an Ising model fit to neural data, it is possible to then test whether the local parameter space of the manifold is flat and uniform, or rather curved and stretched. The above distinction is crucial here; the former type of manifold would imply uniform sensitivity to changes in parameters. Therefore, it would follow that in order to achieve global stability in the presence of fluctuations of local parameters, some form of global homeostatic rule must be in operation, one which takes into account all the activities of units in a group. In contrast, a strongly anisotropic manifold can imply non-uniform sensitivity to changes in parameters, which allows for an alternative scenario of homeostatic compensation – one where control of only sensitive parameters would be sufficient for achieving global stability (however, for that scenario the anisotropy needs to fulfil particular criteria, discussed in the following Section 4.2.4). Thus, understanding the structure of the FIM is of prime interest here.

In order to dissect the local properties of the manifold, the first step is to perform an eigenvector decomposition (i.e. factorization) of the Fisher Information Matrix (Machta et al., 2013) (since the FIM is a real symmetric matrix, it is always diagonalizable and therefore it can always be factorized). This process involves finding a

set of orthonormal vectors $\{v_e\}$ (called eigenvectors) and corresponding scalars $\{\theta_e\}$ (called eigenvalues) such that for each pair the following holds

$$F \times v_e = \theta_e \times v_e. \tag{4.11}$$

In general, the fact that the matrix F in the above equation acts on a set of vectors as a scalar is a very useful property, as it means that in the coordinates defined by the eigenvector basis the action of the matrix corresponds to a simple scaling by an eigenvalue. Since the matrix to be decomposed here is a metric tensor, the resulting vectors will define directions in parameter space 'natural' to the model changes, that is, along the surface of the manifold; and the eigenvalues will determine the magnitude of change along these directions. This procedure can be intuitively understood in geometric terms as a re-parametrization of the space with coordinates induced by the shape of the manifold (see Figure 4.2 for an illustration).

Importantly, the eigendecomposition of FIM automatically gives a handle on the questions posed here. The eigenvalues quantify the importance (i.e. sensitivity to changes) of each direction along the manifold; and the eigenvectors allow to deduce which of the original model parameters contribute to the influential directions. Therefore, if the eigenvalues are of comparable magnitude, it can be inferred that the manifold shape is relatively flat and none of the directions stand out from the rest. In contrast, if the eigenvalues are spread over orders of magnitude, a significant anisotropy of parameter space is present. Such models and their low-eigenvalue insensitive directions have been referred to as 'sloppy' in the literature (Machta et al., 2013), and the high-eigenvalue sensitive directions as 'stiff' (these terms will be used here interchangeably with 'insensitive' and 'sensitive', respectively).

Fisher Information Matrices were computed for each of the fitted models with the use of a custom-written Matlab script. The eigenvector decomposition was then performed with the use of the Matlab function `eig`. As a result, each of the models was characterized by 55 eigenvalues and 55 eigenvectors, each consisting of 55 elements.

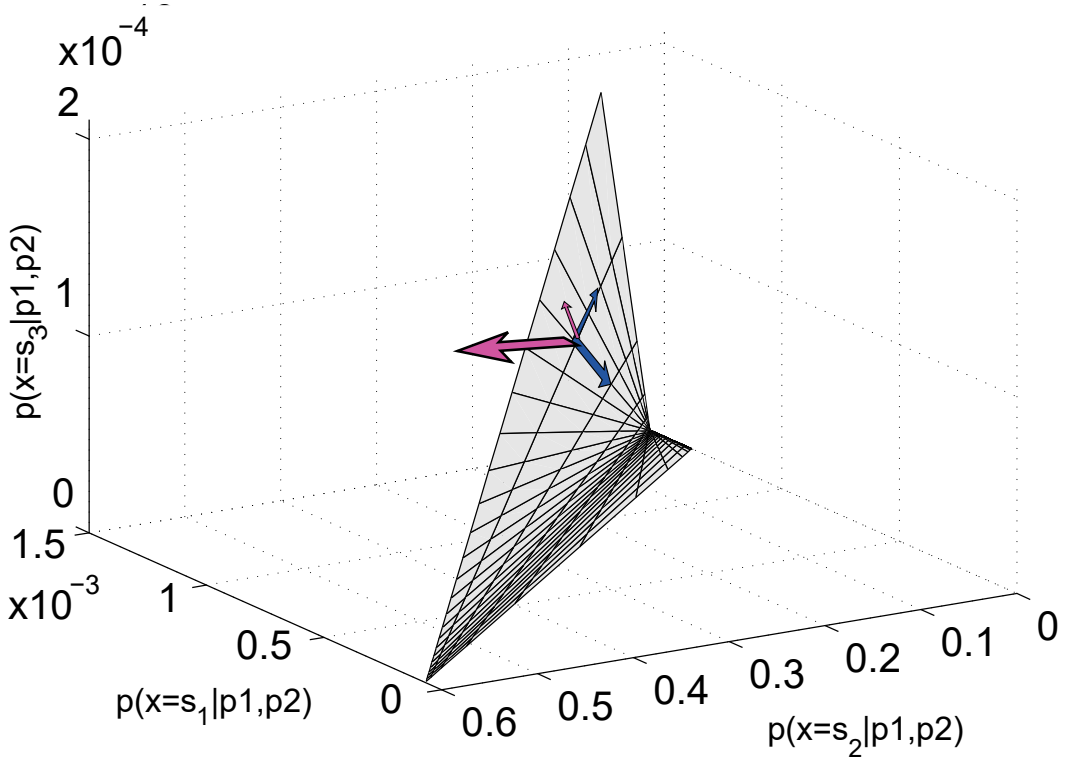

Figure 4.2: Illustration of the concept of locally re-parametrizing the manifold with the dimensions natural to the FIM metric tensor. Blue arrows indicate the coordinates consistent with the parameters of the Ising model. Pink arrows show the dimensions (or technically the 2D projections of dimensions onto the cross-section of the manifold) induced by the Fisher Information Matrix, following the anisotropy of the manifold. The directions on the multi-dimensional manifold are provided by the eigenvectors of the decomposition of the FIM; the eigenvalues inform on the scale of each dimension. The induced vector basis is orthonormal, but here the pink arrows are scaled to illustrate the fact that the models sensitivity to different parameter combinations can vary.

4.2.4 Sloppiness and sparsity analyses

In order to efficiently assess the sloppiness of each of the models, a score was needed for quantifying the statistical dispersion of each eigenvalue distribution. To that end, the Gini coefficient was employed (also known as Gini index), a commonly used and robust measure of sparsity (Hurley and Rickard, 2009). For a distribution of N values θ_i (sorted in ascending order), it is defined as follows

$$G = 1 - 2\sum_{i=1}^{N} \frac{\theta_i}{\sum \theta_i} \left(\frac{N - i + 1/2}{N} \right). \tag{4.12}$$

The values of Gini coefficient range from 0 for the uniform distribution (exhibiting no dispersion) to $1 - 1/N$ for a probability distribution with all mass in one point (i.e. a

maximally sparse N-sized one; referred to thereafter as 'delta' due to its conceptual similarity to the Dirac delta function). For an intuitive understanding, it helps to bring to attention the geometric definition of the index, which is twice the area between the cumulative distribution of the data and that of a uniform vector, as illustrated in Figure 4.3.

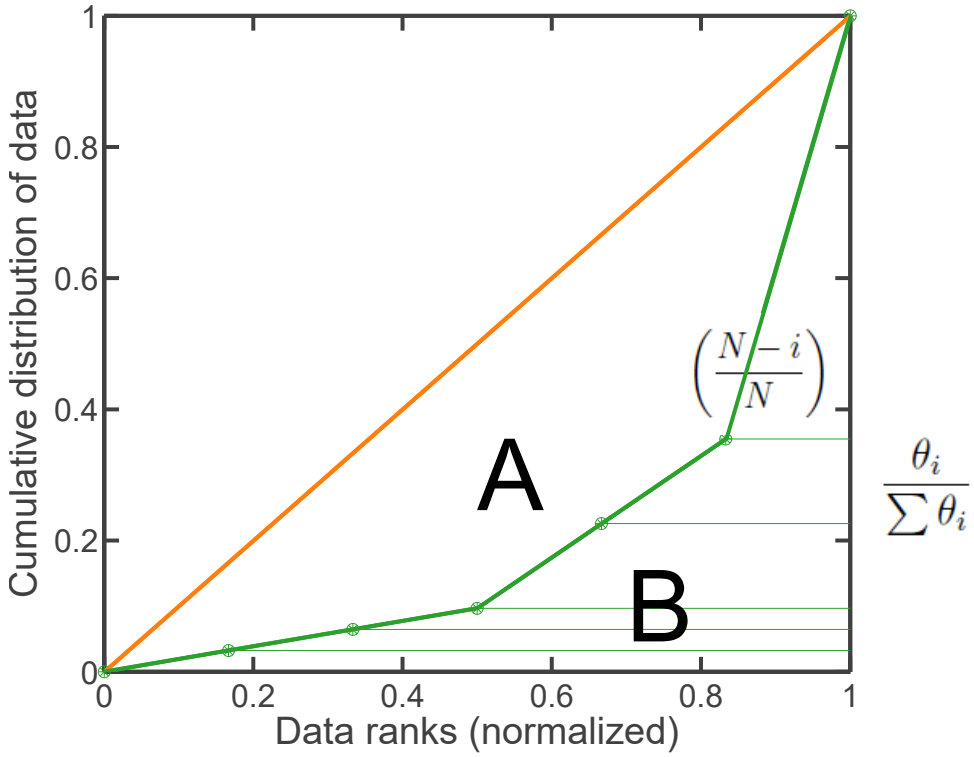

Figure 4.3: Illustration of the geometric interpretation of the Gini coefficient. The coefficient is defined as twice the area between the orange and green curves (cumulative distributions of the uniform and data distributions, respectively), denoted as $2A$. This is equivalent to $1 - 2B$, where B can be easily recognized as the sum appearing in Equation 4.12, being a sum of the areas of the green-bounded trapezes. The more unequally distributed the values of the data are, the larger the A area is.

The index was computed for each eigenvalue distribution obtained, as well as for several well-known reference distributions. The first group of distributions were power laws with exponents α of 1, 1.5 and 2

$$pwr_\alpha(x) \quad = \quad x^{-\alpha},$$

where x is the the rank of eigenvalue, ranging from 1 to 55. The second group of distributions were geometric distributions with associated probabilities p of 0.1, 0.2 and 0.3

$$geo_p(x) \quad = \quad p(1-p)^{x-1},$$

(scaled by the $geo_p(1)$ in order for the geometric sequence to start with 1, for ease of comparison with the eigenvalues). These created geometric sequences with common ratios of 0.9, 0.8 and 0.7. Finally, to provide a non-sparse reference, values exhibiting a linear decrease along the considered range were also used

$$linear(x) \quad = \quad 1 + 1/54 - x/54.$$

In addition to the above mathematical constructs, eigenvalue distributions of simulated data were also employed, for a more natural comparison against model fits. To that end, artificial spike trains were generated of two types: of independent Poisson neurons with firing rates matching those of the data, per each group from baseline recording from culture 1; and of independent Poisson neurons with all equal firing rates (set as an average firing rate of the 10-unit set), per each group from baseline recording from culture 1. These spike trains were then analysed exactly as the original recordings, i.e. were fitted with maximum entropy models, for which then the FIM was calculated and eigendecomposed, to yield the eigenvalue distributions [2].

Establishing the sloppiness of the models was the first step in understanding the structure of the model space and allowed to infer whether there was significant anisotropy in the manifold. However, it needs to be noted here that sloppiness is not equivalent to uneven parameter sensitivity, merely a prerequisite. In order to intuitively understand that, consider for example that the first eigenvector can in principle be a combination of all model parameters with equal weights. Other examples are provided by Machta et al. (2013), who describe how certain microscopic models in physics become sloppy when considered from macroscopic point of view, which is

2 An important side note here is that the quality of these model fits was varied: in some cases independent model had lower Jensen-Shannon divergence than the Ising, in some cases it was the other way around. This is to be expected, since on one hand the data is generated from an independent model, and on the other hand the Ising model has more parameters and is not required to match the first order marginals exactly – which the independent model does. However, since the data was used here for illustrative purposes, the issue was not investigated further.

what allows for a successful dimensionality reduction and effective macroscopic theories. However, in that case the macroscopic parameters are not equivalent to a subset of microscopic ones. On the contrary, none of the individual microscopic parameters are relevant for the macroscopic description (such as in both examples presented by Machta et al. (2013)).

Since the overarching theme of this work was to dissect the interplay between the individual traits of neurons and their group behaviour, it was vital to determine whether in addition to being sloppy, the models were also not uniform in their sensitivity to parameters. In other words, it needed to be tested whether the stiff dimensions themselves are sparse. These two properties of the model, sloppiness and sparsity (understood here as sparsity with respect to individual parameters), are the particular criteria referred to in the beginning of Section 4.2.3, necessary in order to consider a scenario of stability maintenance through control of stiff parameters.

In order to assess the contribution of individual parameters to the sensitivity of the model, rather than looking at each stiff eigenvector in turn, a combined dimension was constructed from the leading directions. This was done by adding together the first several eigenvectors, scaled by the square root of their eigenvalue (recall that the FIM is a metric tensor, entering the equation for path length (Equation 4.3) under square root; therefore the importance of individual eigenvector directions should be measured by the square root of their associated eigenvalues (Daniels et al., 2008; Transtrum et al., 2015)). The number of considered eigenvectors corresponded to the eigenvalue rank at which the residual squared error of FIM reconstruction dropped below 10%, denoted here as $Rk10$; that is, the point at which the combined dimension accounted for 90% of the sensitivity of the model (see illustration of this in Figure 4.4). The 'stiff dimension' constructed in this manner could then be analysed in detail. Furthermore, since the entries in the eigenvectors are essentially coordinates in the space defined by model parameters, the stiff dimension informs on how much each individual parameter contributes to the sensitivity of the model. Therefore, the absolute values of the entries of the combined eigenvector were adopted to serve as sensitivity measures (together forming a sensitivity matrix).

To analyse the stiff dimension, first its sparsity was assessed by the Gini coefficient and subsequently compared to a few informative reference vectors (see Section 4.3.2). Secondly, seeing that a number of groups indeed exhibited considerable sparsity,

the structure of the dimension was examined; specifically, it was tested whether the sensitive parameters tended to be associated mainly with one or two neurons only. This was assessed by comparing the means of the columns of the sensitivity matrix. Finally, it was further analysed whether there was any relation between sensitivity and other individual traits of neurons / neuronal pairs, such as spiking rates.

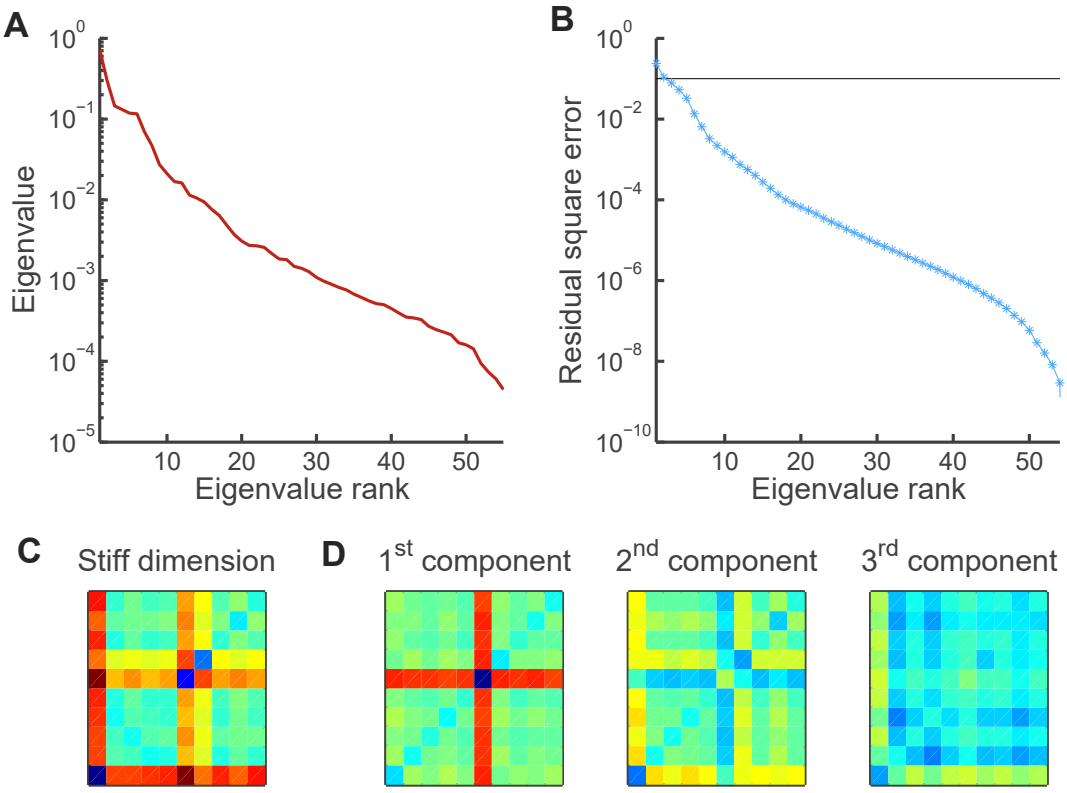

Figure 4.4: Illustration of the construction of the stiff dimension. (A) Example eigenvalue distribution of a single group. (B) The residual square error between the reconstruction from the first r eigendirections and the FIM, plotted as a function of rank r. It is clear that only the first several directions (3 for this particular group) are responsible for the majority of sensitivity in the FIM (here it was arbitrarily chosen to be 90%, black line). (C) The reconstructed stiff dimension, computed as a weighted sum of leading directions. (D) The three components of the stiff dimension, i.e. first three eigenvectors scaled by the corresponding eigenvalue.

4.2.5 Comparing models between time points

Since the pairwise maximum entropy models employed here transpired to be sloppy (as shown in the following Results), it was meaningful to then examine whether this property was exploited by the neurons for stability maintenance in the presence of fluctuations. To that end, the re-parametrization given by the factorization of the

FIM was employed in the following manner. For a given group of units and two consecutive recordings, the model parameters were projected onto the eigenvectors from the baseline recording (i.e. baseline of the two)

$$\psi_i^{t1} = \Lambda^{t1} \cdot v_i^{t1} \tag{4.13}$$

$$\psi_i^{t2} = \Lambda^{t2} \cdot v_i^{t1} \tag{4.14}$$

yielding projection vectors $\Psi_{t1} = \{\psi_i^{t1}\}$ and $\Psi_{t2} = \{\psi_i^{t2}\}$. Due to the common basis, these vectors could then be compared (e.g. as a projection difference $\Delta\Psi = \Psi_{t2} - \Psi_{t2}$) to examine whether there was any relation between the similarity of projections and the rank of associated eigenvalues. The above procedure allowed for assessing what proportion of parametric changes fell into the stiff versus the sloppy dimensions, i.e. whether the changes tended to drive the model up a highly curved direction or rather along the flat subspace.

However, as has been pointed out in Transtrum et al. (2015), sloppiness of parametric models limits the accuracy with which parameters can be estimated. Therefore, a question arises here to what extent do the parameter and projection differences between time points reflect actual changes, and to what extent – the differing accuracy of parameter estimation at different time points. In order to exclude the possibility that any observed relationship stems from the above limitation, additional analysis was conducted for the re-sampled re-fit datasets described in Section 3.2.4. Each of the re-sampled re-fit instances was treated as a new recording, and the average across these provided an estimate of the amount of sloppy and stiff changes expected purely due to fitting inaccuracies.

Another potential factor that could confound results here was the fact that the parameters of the pairwise maximum entropy model come in two forms: individual field and pairwise interaction; and, as shown in Chapter 3, these parameters tend to be distributed differently. Therefore, it could be that one type of parameters exhibits larger changes than the other and these dominate the sloppy subspace, creating a separation in changes with sensitivity. To test against this confounding factor, surrogate models were constructed with field parameters and interaction parameters independently permuted (30 instances of permutation per set of neurons).

4.3 RESULTS

The sloppiness analysis performed here was conceptually similar to other reports (Brown and Sethna, 2003; Daniels et al., 2008), although with certain notable deviations (Machta et al., 2013; Transtrum et al., 2015). These deviations were due to a different modelling strategy adopted in the present work, made possible by the large, detailed datasets and the nature of the system under consideration.

Firstly, since the model used here possesses a parametric form of likelihood, the Fisher Information Matrix was directly calculated from this likelihood (Machta et al., 2013; Transtrum et al., 2015), without the need of postulating Gaussian errors and calculating the Hessian for the χ^2 cost function (Brown and Sethna, 2003; Daniels et al., 2008). Furthermore, all considered parameters were dimensionless, which facilitated the interpretation of tensor factorization (unlike in Brown and Sethna (2003) or some cases considered in Transtrum et al. (2015), where different units of parameters can confound the interpretation of scales of eigenspaces). These differences made it possible to directly identify the shape of the manifold and the directions and dimensions of sloppy and stiff subspaces. Therefore, aside from sloppiness analysis, additional sparsity and sensitivity investigations could be conducted upon the stiff subspace of the FIM decomposition. Secondly, obtained re-parametrization of the manifold was exploited to directly test which groups changed in time.

The FIM was calculated and subsequently eigendecomposed for each of the fitted Ising models. That is, for each pair of consecutive recordings and each of the chosen groups of neurons, re-parametrization of the model space was obtained for the first and second recording.

First, to illustrate the usefulness of the Fisher Information Matrix and its decomposition for understanding the models, results for a single group of neurons from culture 1 are presented in Figure 4.5. Already a visual examination of the 55x55 FIM (panel B of the figure) reveals that the tensor matrix is sparse but structured (note that most entries are 0 or very nearly so, and the highest-value entries are regularly distributed). Therefore it seems likely that the model will exhibit signatures of sloppiness; furthermore, it also seems likely that the sensitive dimensions themselves might be non-uniform and structured.

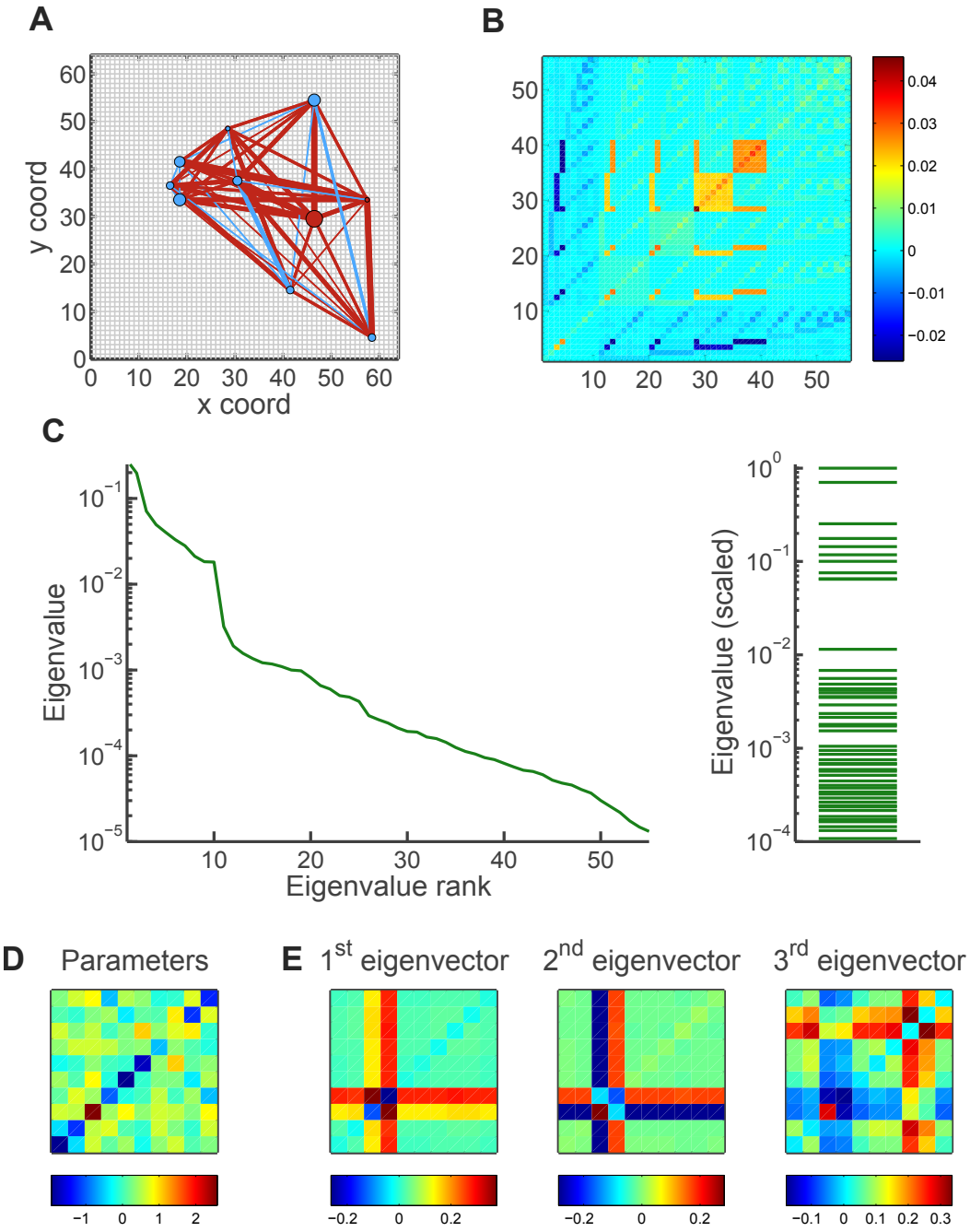

Figure 4.5: Illustration of the Fisher Information Matrix analysis. (A) Parameters of the fitted model of a single group of 10 neurons at 0h (i.e. the baseline recording), culture 1, pictured over the MEA layout; line widths and circle diameters are proportional to corresponding parameter values, interactions and fields; blue indicates a negative sign and maroon a positive one. (B) The Fisher Information Matrix of the fitted model; many of the entries are close to zero (implying flat subspaces) but some specific combinations of parameters are sensitive to changes. (C) Eigenvalues of the eigenvector decomposition of the FIM from panel B. Values are spread over several decades (note the log ordinate scale) with most eigenvalues two orders of magnitude smaller than the first one, implying that most of the directions in parameter space have little impact on the model. (D) Parameters of the fitted model in the form of a matrix, with color-coded values; fields are plotted on the diagonal, interactions off-diagonal. There is little apparent structure in the parameter matrix. (E) The first three eigenvectors of the FIM decomposition. In this re-parametrization it becomes clear that only a subset of the 55 parameters are highly influential over the behaviour of the fitted Ising model.

Indeed, the distribution of eigenvalues is sparse (Gini coefficient 0.92), with the values spread over several decades and the majority of directions having little influence over the FIM (see Figure 4.5C). Clearly, the first two dimensions are dominant in this group of neurons, and dimensions beyond the 10^{th} are of little consequence. Interestingly, the first two eigenvectors exhibit noticeable regularities, in contrast to the parameters themselves, which appear not very structured (compare panels D and E of Figure 4.5). Many of the entries in both matrices are close to zero, and the highest values cluster into two bands.

This indicates that the most influential directions along the hyper-surface of the manifold are in fact a combination of only a subset of parameters. Here, in the case of both the principal eigenvector and the second-ranked, it is the fields and interactions of the 3^{rd} and 4^{th} units that determine the two stiffest directions. Therefore, it transpires that for the particular chosen group here only two neurons and their interactions largely control the group behaviour.

Since the model appears sloppy, it becomes then meaningful to examine what happens over time to the projections of parameters onto the eigenvectors. In that manner, it can be tested whether the changes that occur between recording points take place predominantly along sensitive or, rather, insensitive directions. In Figure 4.6A the individual elements of projection vector from the baseline recording are plotted against those from the recording at 2h. For ease of visual comparison, each point is both color-coded and plotted along the z-axis, using the rank of associated eigenvalue. In addition, panel B of the figure shows the projection of these data points onto the horizontal plane, to provide a complementary point of view and verify the relation of the points with respect to the diagonal.

What transpires from this picture is that in the first few dimensions there were hardly any changes between recording points: the darkest blue points in Figure 4.6 lie along the diagonal, which means that the leading entries in the projection vectors were almost the same. However, with increasing rank of dimension, the projections funnel out (see panel A) and changes become apparent – lighter blue points begin falling off the diagonal, and some of the largest differences can be observed for the yellow to red hues. Recall that the after the first few eigenvalues, the relevance of subsequent dimensions is at least an order of magnitude smaller than the principal ones. Thus it appears that for the presented group of neurons, although parameters

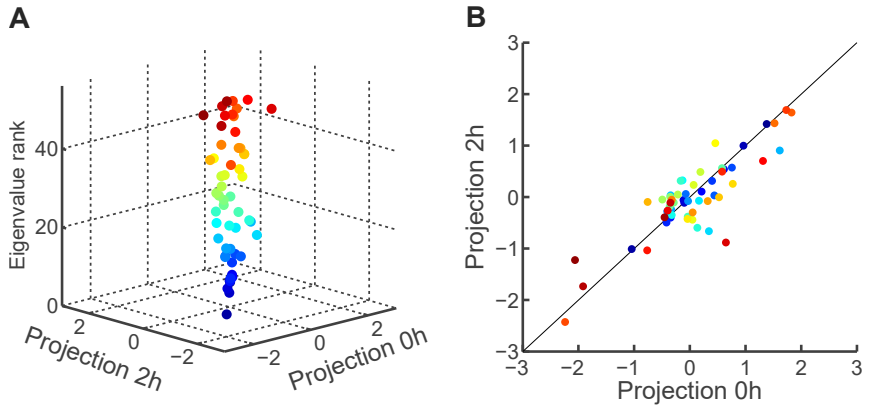

Figure 4.6: Illustration of the comparison of projections between time points. (A) Plotted along the x-axis are individual entries of the vector Ψ_{0h}, that is, a projection of the parameters at 0h onto the eigenvectors of the FIM at 0h. Along the y-axis, corresponding entries of the vector Ψ_{2h}, that is, a projection of the parameters from second recording onto the eigenvectors from the first recording (in order for the two vectors to be comparable they need to be expressed in the same coordinate system). Along the z-axis is the rank of eigenvalue associated with each eigenvector. The vector entries are also color-coded according to their eigenvalue rank, with the navy blue corresponding to the first, stiffest eigenvector, going through the palette (blue, green, yellow, red) towards the sloppiest ones. (B) Same data points as in (A), projected onto the horizontal plane, with a diagonal line plotted for visual aid.

do change between recording points, these changes are largely confined to sloppy dimensions.

This is a particularly interesting result, as it suggests a possible mechanism of global homeostasis in groups of neurons, without a concurrent loss of individual plasticity. Namely, in sloppy and sparse systems, stable collective behaviour can be ensured by a minority of stiff parameters, with sloppy subspaces allowed to fluctuate and, for example, undergo learning-related plastic changes. In the following subsections all the fitted models from each pair of recordings are tested for signatures of sloppiness and sparsity, and the support for the above hypothesis is examined.

4.3.1 Evaluation of sloppiness of the maximum entropy models

The first step of the analysis was to assess whether all of the models exhibited sloppy behaviour. As a starting point, for visual assessment, in Figure 4.7A the distributions of eigenvalues are plotted for each set of neurons from the baseline recording in culture 1. In all cases the eigenvalues decrease substantially with each rank and the majority are orders of magnitude smaller than the leading ones; however, there is a degree of variability between groups of neurons. The variability is most pronounced

in the area of the first few ranks, where the decreases in eigenvalues exhibit different slopes. This behaviour can be interpreted as somewhat varying degrees of sloppiness, which is understandable: some models reside on very stretched and curved areas of the manifold, while others in relatively flatter ones.

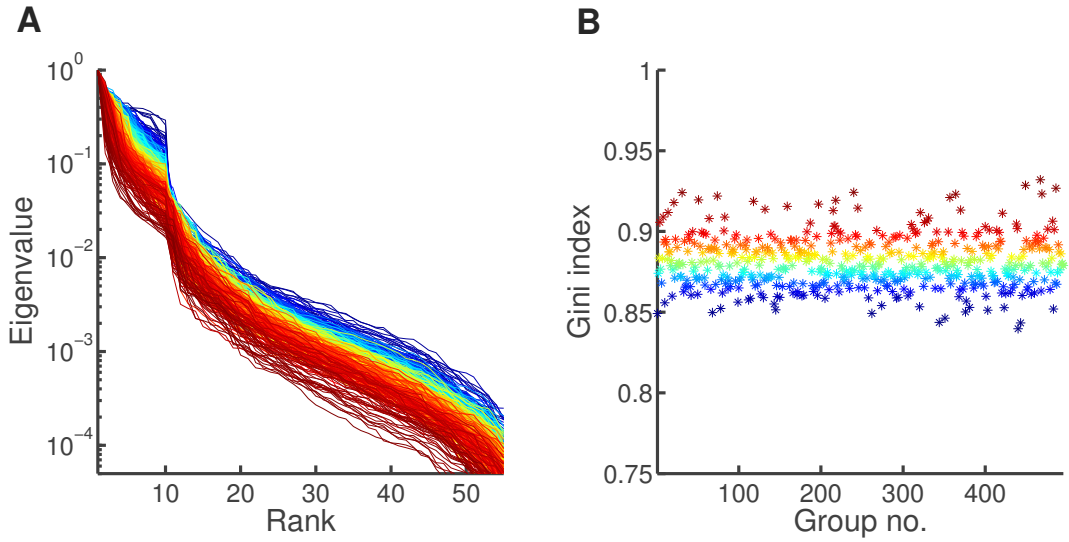

Figure 4.7: The distributions of eigenvalues and corresponding Gini coefficients from the baseline recording in culture 1. (A) Eigenvalue plotted as a function of its rank for each model fit; eigenvalues are scaled to the first one, for ease of comparison; lines are color-coded according to the magnitude of their Gini coefficient, see panel B. While there is some variability between sets of neurons considered, in all cases the eigenvalues exhibit signatures of sloppiness, i.e. values spread over several decades, with most at least two orders of magnitude lower than the first. (B) Gini coefficients calculated for each of the distributions from panel A.

However, since sloppiness is not a stringently defined property, there is no associated measure that could be used to qualify the above statement or compare against other studies. Here, in order to provide a somewhat more systematic approach, the Gini coefficient is introduced as a measure of statistical dispersion of the eigenvalues. In panel B of Figure 4.7, the values of the index are plotted for each eigenvalue distribution from panel A. It can be observed that these fall into a relatively narrow range, clustering between 0.85 and 0.9, which suggests that indeed for all considered models the eigenvalue distributions were sparse (recall that low Gini index is associated with uniform distribution, while values close to 1 indicate very sparse distributions).

To further quantify sloppiness, in Figure 4.8 the average eigenvalue behaviour across all models is compared to other well-known sparse distributions: the geometric distribution (where discrete probabilities form a geometric sequence; this has been highlighted by other authors as a signature of sloppiness (Machta et al., 2013;

Transtrum et al., 2015)); and the power law distribution (associated with scale-free behaviour). Also for reference are plotted non-sparse distributions: uniform, and linear.

As can be seen in panel A of Figure 4.8, the eigenvalues prove to exhibit interesting behaviour. Along the first several ranks, the distribution closely resembles a power law with exponent -1. However, it then becomes steeper, along a further few ranks reminiscent of a geometric sequence. Then in turn it again becomes less steep, to roughly follow a power law with exponent -2. The source of this irregularity could lie in the finite size of the modelled system, although this remains unclear at this point. However, the above similarities leave no doubt that the eigenvalue distributions are sparse and the corresponding models sloppy. It should be noted here that the above comparisons were for illustrative reference only, and these distributions were not rigorously fitted to the distribution of eigenvalues. While the similarities to power laws are intriguing, especially in light of recent interest in criticality in neuroscience (Shew et al., 2011; Beggs and Timme, 2012), investigation into this area was beyond the scope of this work.

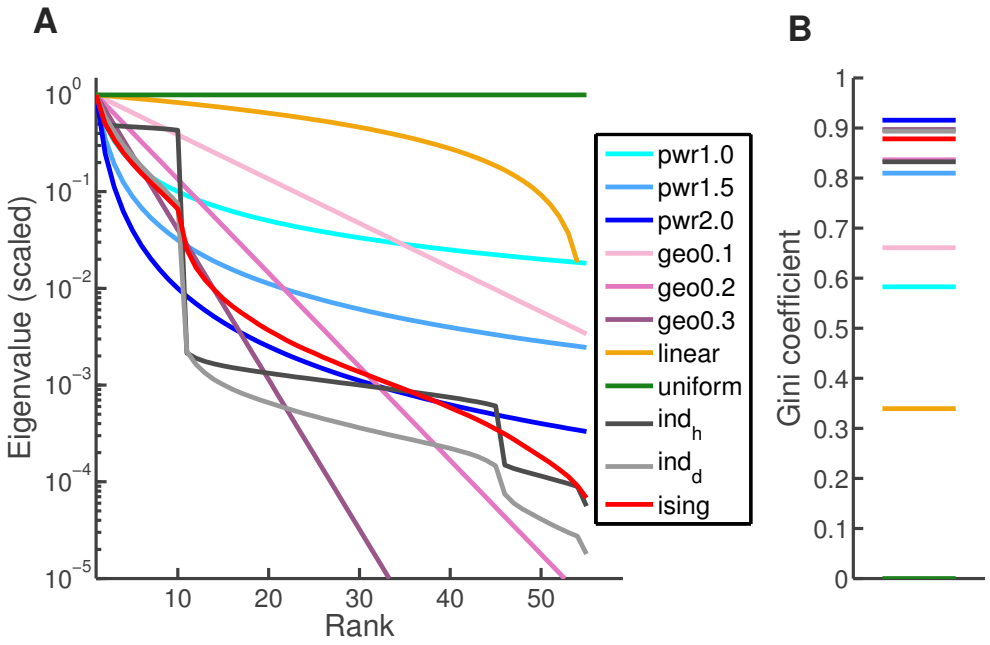

Figure 4.8: Comparison of the dispersion of eigenvalues to reference distributions. (A) Average scaled eigenvalue plotted as a function of its rank (average across all sets from culture 1, and scaled to the first one) and reference distributions: three power-law distributions with varying exponents; three geometric distributions with varying probability parameter; a linear and a uniform distribution. (B) Gini coefficients calculated for each of the distributions from panel A; same color-code as in panel A.

Comparisons to mathematically specified distributions can be useful in generating an intuition about sloppiness and as a benchmark against other studies; however, it is also instructive to compare eigenvalue distributions obtained here against those of other models. Therefore, for a realistic comparison against models expected to exhibit different profiles of sloppiness, simulated Poisson spike trains of two types were also analysed here (see Section 4.2.4). First, let us examine the surrogate data in which the firing rates of all units were set to the same value. Denoted in Figure 4.8 with dark grey, the average eigenvalue distribution for models of this type has a particular shape, with a sharp decline after rank 10 separating two relatively flat parts of the curve. The initial flattened portion results from the fact that in these models each of the units is in principle equally influential on the behaviour of the group (barring the differentiation resulting from the sampling noise). This, together with the fact that the remaining 45 eigenvalues are drastically lower, might appear to suggest that under this scenario the only truly sensitive parameters were the fields associated with each unit, and the leading eigenvectors are each associated with those parameters; and the interaction parameters then fall into the insignificant portion of the eigenspectrum beyond the 10th rank. Note that all of the model-derived eigenvalue distributions feature a deflection at rank 10, and so it might seem compelling to assume this pattern holds universally, separating 10 field parameters and 45 interaction parameters. This is, though, not true in fact for either of the models and the picture here is more subtle.

In the next subsection, the stiff dimension of the biological Ising models will be analysed thoroughly, but the matter can be already intuitively grasped by considering the nature of interactions in the simulated Poisson data and the notion behind FIM eigendecomposition. The interactions of Ising models for Poisson spikes need to be small and insignificant, in order to well capture the independence of individual spike trains. However, insignificant does not equate insensitive – on the contrary, these parameters are quite sensitive, since any changes in those will throw off model predictions not just for one unit (as a change in field parameter would) but for both connected units, breaking the independence assumption. Now recall that the idea behind eigendecomposition is to identify the directions in parameter space that result in most drastic changes in the model. For a model of homogeneous Poisson spikes, what would result in largest deviation from its original predictions is e.g. increasing in unison most of its interactions, rather than just changing a single field parame-

ter. And indeed, the leading eigenvectors are not each associated with a single field parameter, but rather with a combination of many parameters.

The fact that the Ising models of biological data are more sloppy than those of the homogenous Poisson spikes suggests therefore that for the dissociated cultures there exists some structure in sensitivity to changes. What can then be inferred from the comparison against the data-matched Poisson spikes? In this case the sloppiness was higher than that of the biological Ising models, not by a large margin, but significantly so. The average eigenvalue distribution (denoted with pale grey in Figure 4.8) follows closely the one obtained from the data up till rank 10, to then sharply drop. This, again, might appear to follow from the fact that each eigenvalue would correspond to a single units' field; but considering the fact that it is important for the interactions to remain low and insignificant, an alternative explanation emerges: it is rather that all the parameters of a single unit need to remain stable; and the characteristic 10th rank break-off is due to the fact that each group consists of a collection of units with varying importance for the group. Thus, e.g. a strongly firing unit in the independent Poisson simulation has interactions that are more sensitive than those of other units.

The fact that the biological Ising and data-matched Poisson exhibit very similar eigenvalues up till rank 10, and that the main difference in sloppiness profile occurs beyond this point, might raise doubts as to whether interactions are in fact all that important in modelling the data; e.g., it could be that the stiffest eigenvector corresponds to a highly active but isolated unit. To answer this question, the structure of the stiff dimension needs to be examined in detail, along with corresponding parameters; this is the purpose of the next subsection in this chapter.

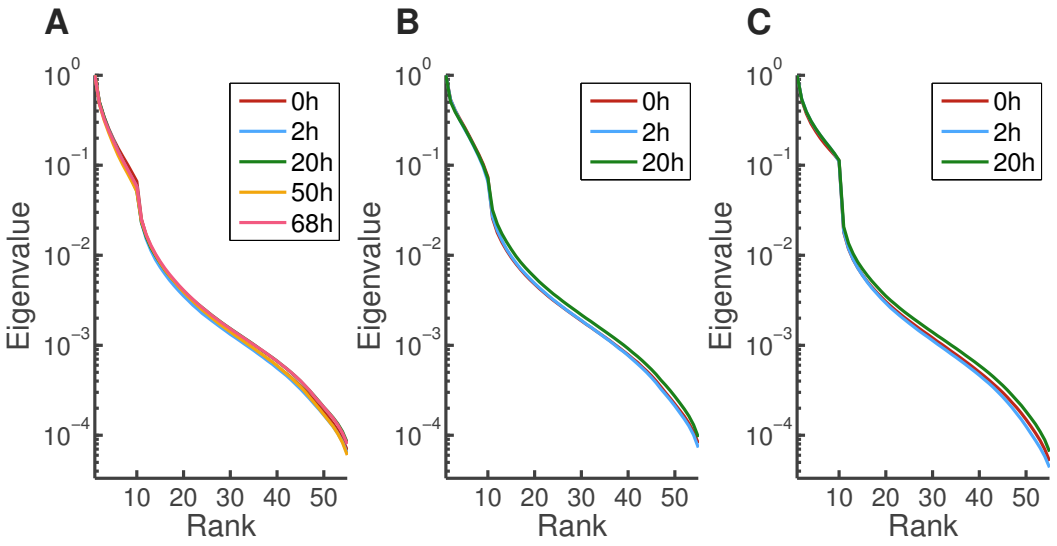

Figure 4.9: Sloppiness of the models across cultures and recordings. (A) Average eigenvalue distribution (scaled by the first one) from all sets across recordings in culture 1; reported only baseline recordings from each pairwise comparison, as these were representative. (B-C) Same for cultures 2 and 3, respectively.

The above results reported in detail for culture 1 held across all recordings and all cultures. As summarized in Figure 4.9, the characteristic sparse eigenvalue distribution, with behaviour roughly approximated by a mixture of power law and a geometric distribution, was present across the datasets. Again, as in previous chapters, slight differences between cultures can be observed, such as a sharper inflection in the curves in Figure 4.9C, occurring at rank 10, rather than approx. 13 in panel A, or a nearly smooth transition in panel B. These subtle variations might be related to the inhomogeneities between non-sister cultures discussed in Chapter 2. It is tempting to speculate here whether these discrepancies are related to clustering properties of each of the cultures. Recall that culture 3 was the one with most prominent clustering; since here it also appears to exhibit the sharpest distinction between sloppy and stiff eigenvalues, it might be due to higher prevalence of hubs of neuronal activity, distinct from 'regular' un-influential neurons. It should be noted, however, that no formal analysis of clustering was performed here, and therefore the above is a speculation only.

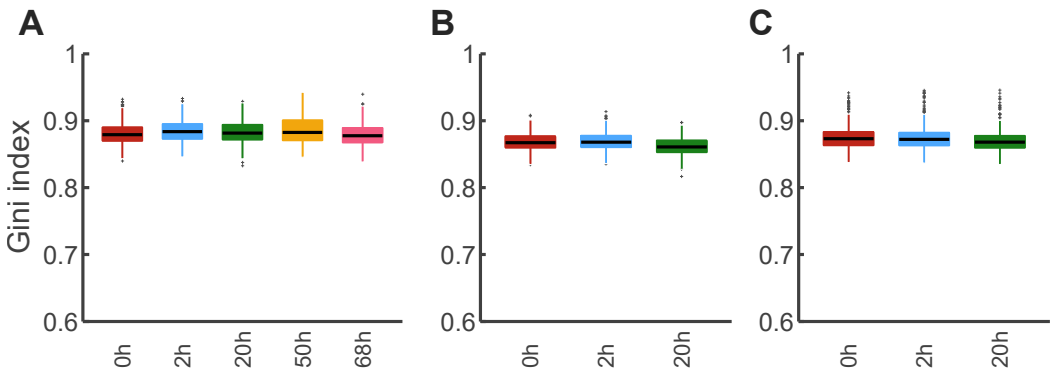

Figure 4.10: Sloppiness of the models across cultures and recordings. (A) Gini coefficients of the eigenvalue distributions obtained from model FIMs in culture 1; reported only baseline recordings from each pairwise comparison, as the Gini coefficients shown here are representative. (B-C) Same for cultures 2 and 3, respectively.

Sloppiness was apparent also in the behaviour of eigenvalues of each individual model. The Gini coefficients were consistently close to 0.9, and varied little between sets of neurons or cultures, as shown in Figure 4.10.

4.3.2 Evaluation of the sparsity of the stiff dimension

Having established that the pairwise maximum entropy models fitted to neural data exhibited sloppy behaviour, the anisotropy of the parameter space could then be tested in more detail. To that end, stiff dimensions were constructed as explained in Section 4.2.4, with entries providing a proxy for individual parameter sensitivity (specifically, the absolute value of vector entries, since sensitivity is directionless).

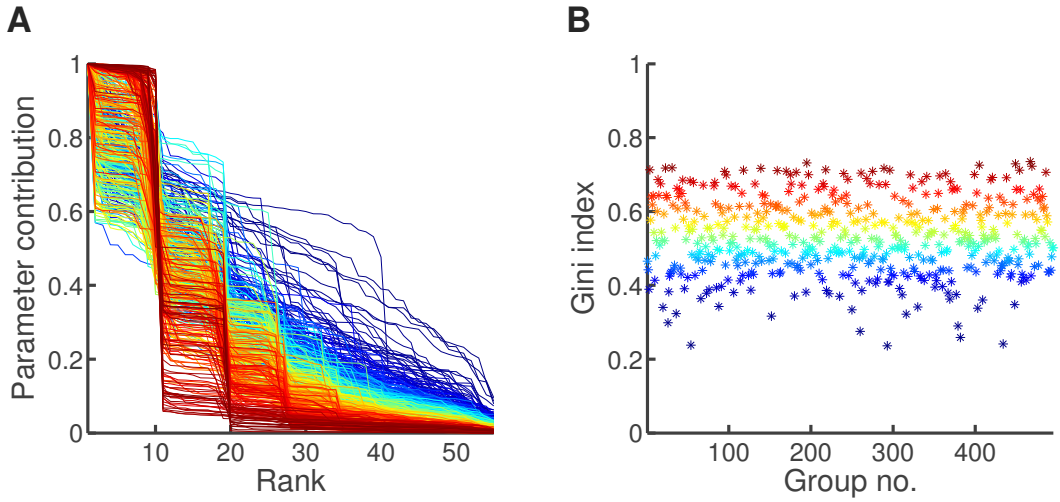

Figure 4.11: The distributions of individual parameter contributions towards the stiff dimension and the corresponding Gini coefficients for the baseline recording in culture 1. (A) Ranked and scaled magnitudes of entries in the stiff dimension (a weighted sum of the leading eigenvectors), i.e. essentially a distribution of sensitivities of individual parameters for each set of neurons; lines are color-coded according to the magnitude of their Gini coefficient, see panel B. There is considerable variability between groups, with only a subset exhibiting clear signs of recognizable structure. (B) Gini coefficients calculated for each of the distributions from panel A.

First, for visual assessment, in Figure 4.11A plotted are the distributions of sensitivities for each set of neurons from the baseline recording in culture 1. Several things are apparent in this picture: considerable variability between groups; intermediate level of sparsity (note that the ordinate scale is linear and compare with Figure 4.7A); and occurrence of particular step-shaped distributions in a number of groups. The latter is indicative of underlying structure in the stiff dimension, possibly a confinement of sensitivities to one or two neurons (recall that the example in Figure 4.5 exhibited such structure). Examining the Gini coefficients in Figure 4.11B reveals that the sparsity is both significantly lower than that of eigenvalues and much more varied across sets of neurons. Spanning the range of approximately 0.3-0.7, it thus places somewhere between a linear and exponential or geometrical decay (see Figure 4.8B for reference Gini indexes).

In order to interpret the above results, again it is useful to examine some reference distributions and their Gini coefficients (Figure 4.12). However, since the focus of interest here is a vector defining a direction in parameter space, the considered examples need to be different than in Section 4.3.1. In particular, the chosen distributions were constructed from other types of vectors that might be pertinent in the context of

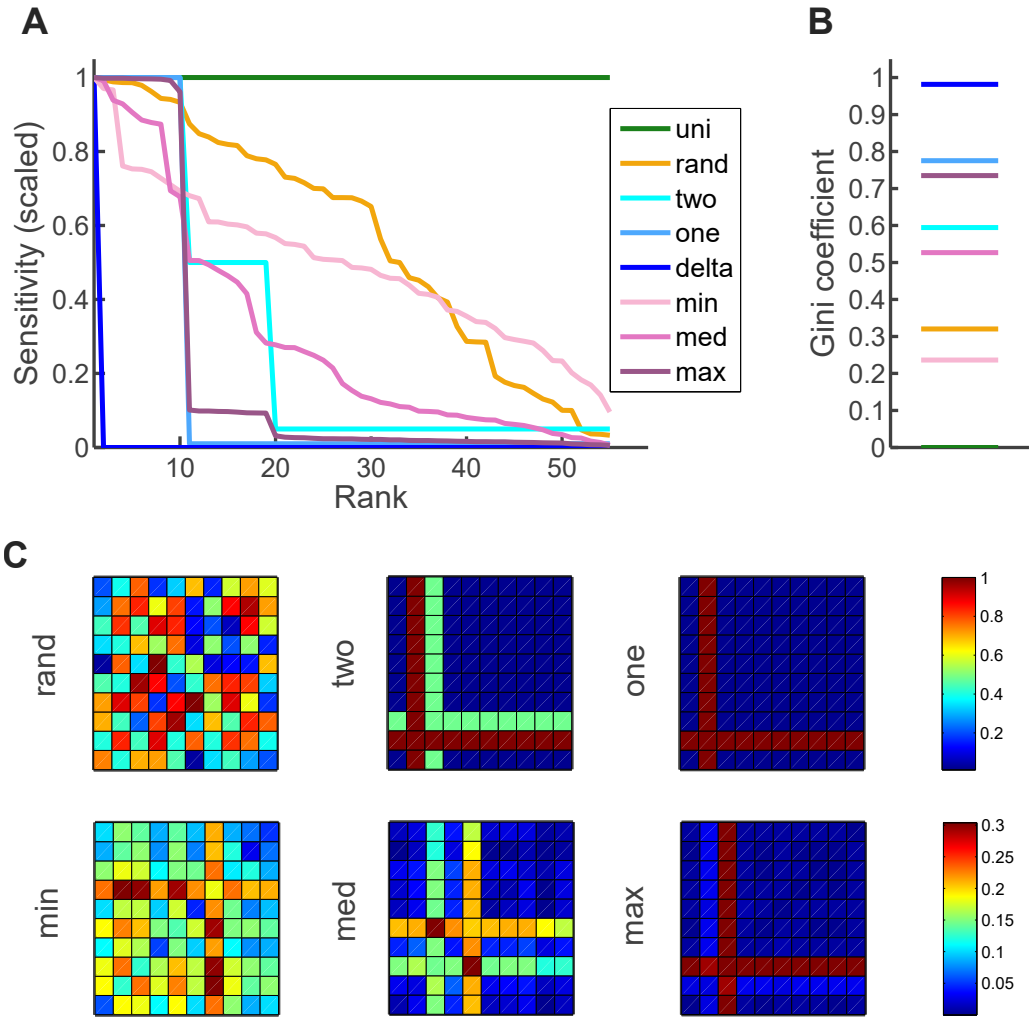

Figure 4.12: Comparison of the stiff dimension sparsity to chosen reference vectors. (A) Scaled sensitivity (the absolute value of the stiff dimension vector entry, scaled to the largest one) of three groups of neurons from culture 1, corresponding to the minimum, median and maximum Gini coefficients. Also plotted distributions of entries of reference vectors: uniform; random; corresponding to a single direction (delta); and vectors corresponding to one and two neurons' parameters (see subplot C for visual reference). (B) Gini coefficients calculated for each of the vectors from panel A; same color-code as in panel A. (C) Illustration of the reference vectors and the stiff dimensions pictured in A, as labelled (shown in matrix form for convenience; recall that model has 55 parameters and interactions are symmetric). Note that the sparsest groups resemble the 'one' reference, i.e. all parameters of a single neuron determine the stiff dimension; and the medium-sparse groups resemble the 'two' reference, i.e. two neurons determine the stiff dimension.

mapping parameter space: a vector determined by a single parameter ('delta', see Section 4.2.4); a vector with random entries; vector with 10 entries set to 1, and remaining ones to 0.01, corresponding to a situation where one neuron and all its parameters define the stiff dimension, with remaining parameters small enough to be inconsequential ('one', see Figure 4.12C); and a vector with 10 entries set to 1, 9 entries set to 0.5 and the rest to 0.05, corresponding to a situation where two neurons and their parameters define the stiff dimension – however, with unequal contributions ('two', see Figure 4.12C).

What transpires from this comparison is that the Gini indices of examined sensitivities span roughly the range between a random vector and that of a single-neuron dimension. That is, while none of the groups were so sparse as to be controlled by a single or very few parameters, they were also far from a uniform sensitivity. The sparsest of pictured groups (labelled 'max' in Figure 4.12) exhibits the characteristic step-wise distribution of stiff dimension entries, with exactly 10 vastly larger than the rest. Furthermore, as shown in panel C of Figure 4.12, these entries are all associated with a single unit. An interesting case is the 'med' example, which although similar in sparsity to the 'two' reference vector, proves to exhibit a more complicated structure. Although in both cases it is the contributions of two neurons that dominate the vector (see Figure 4.12), there is a further dissociation in the distribution of sensitivities of the 'med' group, brought on by the fact that the interaction parameter between those two units stands out from the rest. Finally, the least sparse group from culture 1 has a Gini index in fact lower than a random vector, although remnants of structure can still be discerned in the matrix (notice that in the 'min' panel of Figure 4.12C the highest values still cluster into a band corresponding to a single unit).

In order to better understand the structure of the stiff dimension suggested by the step-wise sensitivity distribution (Figure 4.11A) and what it implies about the possible role of the neurons in the group, in Figure 4.13 a range of representative examples are presented in more detail.

In the top three panels, the characteristic single-step distribution is taken under consideration. Although it is tempting to assume that the upper 'step' corresponds to a single unit, since the sensitivities are ranked according to magnitude, it cannot be excluded without a further check that these might be a combination of parameters from different units. To test against this possibility, Figure 4.13B plots sensitivities that

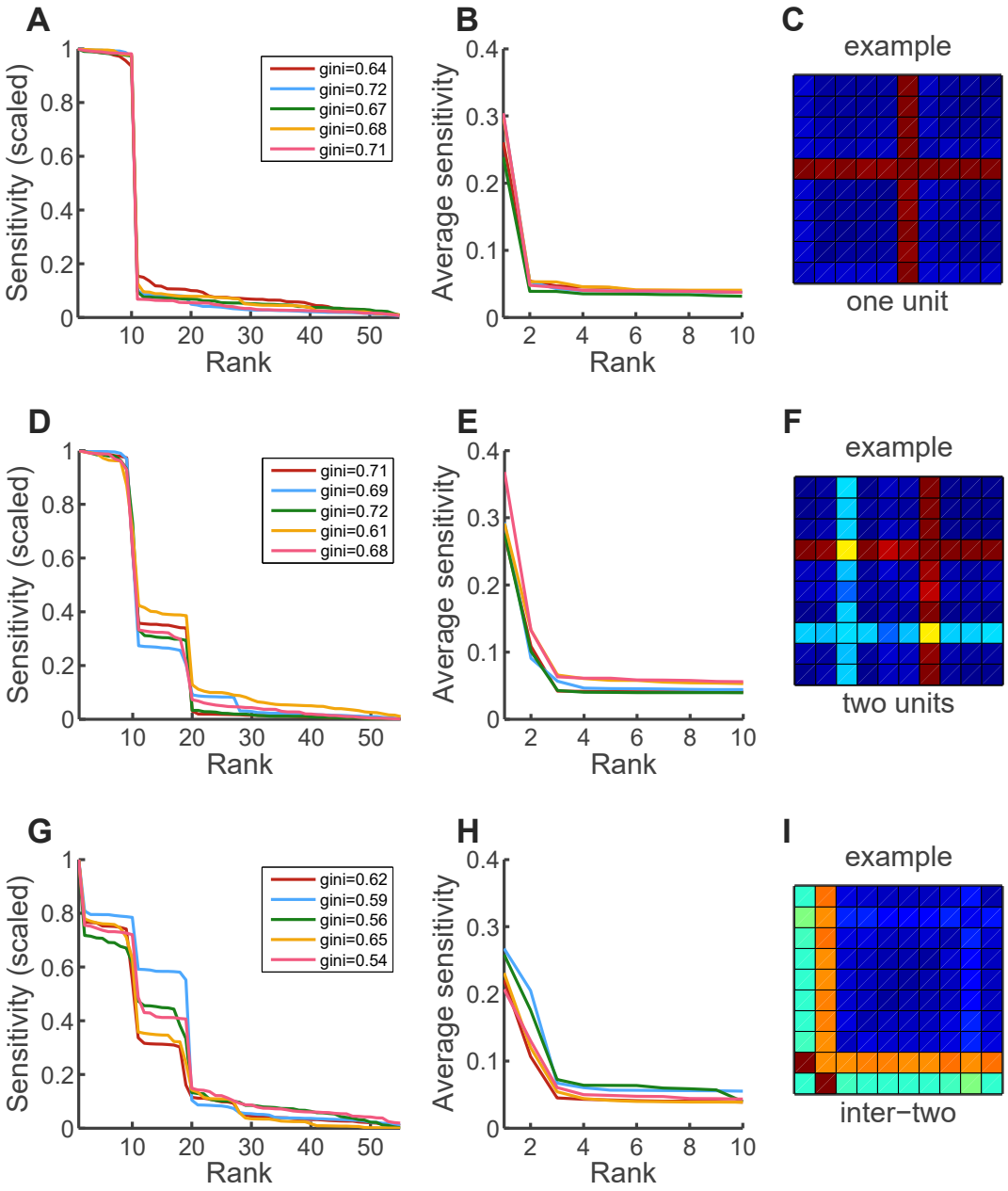

Figure 4.13: Representative examples of different group structures. (A) Distributions of parameter sensitivities for 5 representative groups of neurons with a similar signature shape, indicative of a single-unit structure of the stiff dimension. (B) Averaged sensitivity of each unit, ranked by magnitude. Note a definite dominance of sensitivity of a single unit, confirming the intuition from panel A. (C) Sensitivity matrix of one of the groups presented in A and B, as a representative example (the same pattern was present across groups). (D) Distributions of parameter sensitivities for 5 representative groups of neurons with a similar signature shape, here (in contrast to A) indicative of a two-unit structure of the stiff dimension. (E) As in B, averaged sensitivities across neurons. Note that here it is the first two units that stand out, again confirming the intuition from the distributions panel, D. (F) As in C, a representative sensitivity matrix corresponding to previous panels, D and E. (G) Distributions of pattern sensitivities for 5 representative groups of neurons with a similar signature shape, here (in contrast to A and D) indicative of single parameter dominance, followed by two units. (E) Analogous to B and E, averaged sensitivities per neuron. Note that again only two units stand out from the rest. (F) Analogously to C and F, a representative sensitivity matrix. Interestingly, the most sensitive parameter is the interaction between the two most sensitive units.

were averaged separately for each neuron, and then ranked. This clearly shows here that for all considered examples it was a single unit dominating the stiff dimension; and (as can be seen from panel A), it was all of its parameters that were sensitive, i.e. both the field and the interactions. For a visual check, panel C of Figure 4.13 shows a representative matrix of this type of structure. It appears thus that for those groups, there was a single dominant unit, perhaps akin to a what is considered a hub neuron in small-world networks; however, since there are alternative confounding interpretations, further analysis will be presented in several paragraphs.

The second row of panels of Figure 4.13 deals with the 'two-step' type of distribution encountered in Figure 4.11A. Again, it is compelling to assume that the group might be mainly under the influence of two neurons. And indeed, panel E of Figure 4.13 confirms that scenario: although again the first ranked unit is dominant, here, in contrast to panel B, also the second unit stands out above the rest. A look at the sensitivity matrix in panel F visualises this for the reader, showing clearly two bands, one however of higher magnitude than the other. Again it is tempting to see this as evidence of a particular network structure, where two hub neurons happen to be chosen for the ensemble, and thus both are found to be influential. An analogy to the rich-club network hypothesis (Buzsáki and Mizuseki, 2014) also comes to mind, since the contributions of the two neurons are unequal – namely, the first one appears as being the 'richer', i.e. a better-connected one.

Finally, the bottom panels of Figure 4.13 deal with an interesting case brought up in Figure 4.12E. As can be seen in panel G, this time the distribution drops steeply after a single parameter (rather than 10, as in panels A and D), to only then assume again the regular step pattern. This single most sensitive parameter is the interaction between the two most sensitive neurons of the group, as can be confirmed from examining the average sensitivities, where again only two units stand out from the rest (if this singular parameter belonged to a different unit, then also the third ranked unit would exhibit a noticeable increase above the remaining neurons). Continuing the intuition from previous paragraphs, this could be interpreted as two 'rich' neurons controlling the group; since they are, however, comparably well-connected, it becomes crucial what the interaction between the two of them is, as changing it might e.g. disturb the balance in the group.

In contrast, groups in the lowest percentiles of Gini index range show only rudimentary structure in the stiff dimension, with less separation in contributions from individual neurons. Presumably, the random sampling resulted in a relatively balanced ensemble in terms of connectedness and activity levels; this is not surprising that a number of such groups would be sampled, since one of the tenets of rich club hypothesis is that the highly influential units constitute the minority of neural population. Note, however, that very few groups exhibit values of Gini coefficient lower than 0.4 (Figure 4.11B).

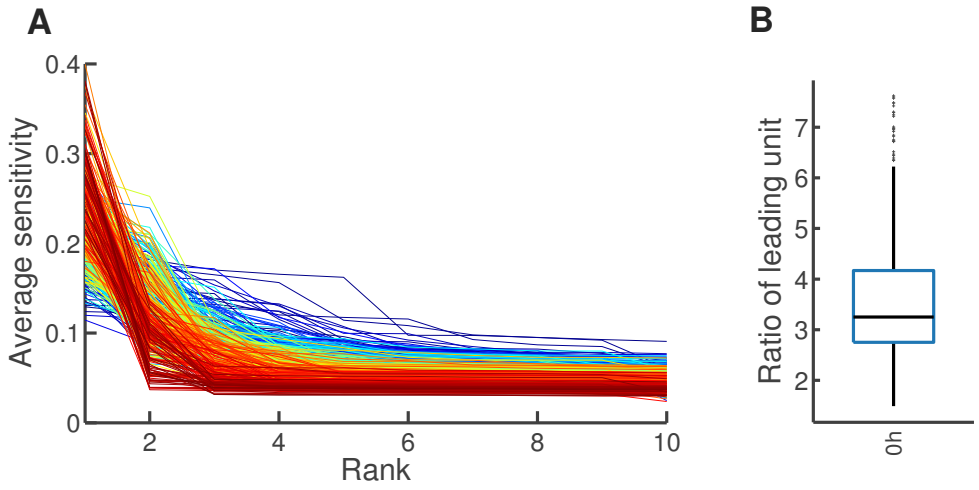

Figure 4.14: Average parameter sensitivity per neuron in culture 1. (A) Averaged for each unit, contributions of individual parameters towards the stiff dimension, ranked by magnitude; lines are color-coded according to the sparsity of the stiff dimension. Note that for the groups with sparsest stiff dimension the sensitivity is clearly confined to a single neuron. (B) The distribution (across groups of neurons) of ratios of the average sensitivity of the leading unit versus average sensitivities of remaining neurons.

To further validate that the above-mentioned observations of sensitivity confinement to one/few units held for the majority of groups, the average values of sensitivity per each neuron were calculated for each of the groups in culture 1. In Figure 4.14, the mean contributions towards the stiff dimension of each unit are plotted, sorted by magnitude. For most of the groups this distribution exhibits a clear peak at the origin, and plateaus from the 2^{rd} / 3^{rd} unit. This confirms that in the step-wise pattern observed in Figure 4.11A, indeed the 'steps' correspond to parameters of a single neuron, as opposed to a random combination of 10 parameters, or e.g. all the field parameters of a group. Ruling out the latter example is of particular importance here, since a stiff dimension composed primarily of field parameters would imply that the interactions of the Ising model are inconsequential.

To summarize the above for each of the groups, the ratio of the leading unit average sensitivity to the average sensitivity of remaining units was then computed. As can be seen in Figure 4.14B, for almost 75% of groups this ratio was higher than 3, and for nearly all more than 2. Thus it transpires that across the examined groups, the parameters of a single unit were at least twice as sensitive as those of the other neurons.

In order to better understand what the results reported thus far imply about the structure of groups and relationships between units, first, in Figure 4.15 the average sensitivity is plotted against the neuron's field parameter and its firing rate. What becomes clear here is that the most sensitive units were also the ones with the highest field parameters and firing rates. Thus it would appear that the sensitive dimensions were dominated by highly active neurons. Interestingly, Figure 4.15B bears certain similarity to Figure 3.8A, with the nonlinear preservation of rank and characteristic vertical bands.

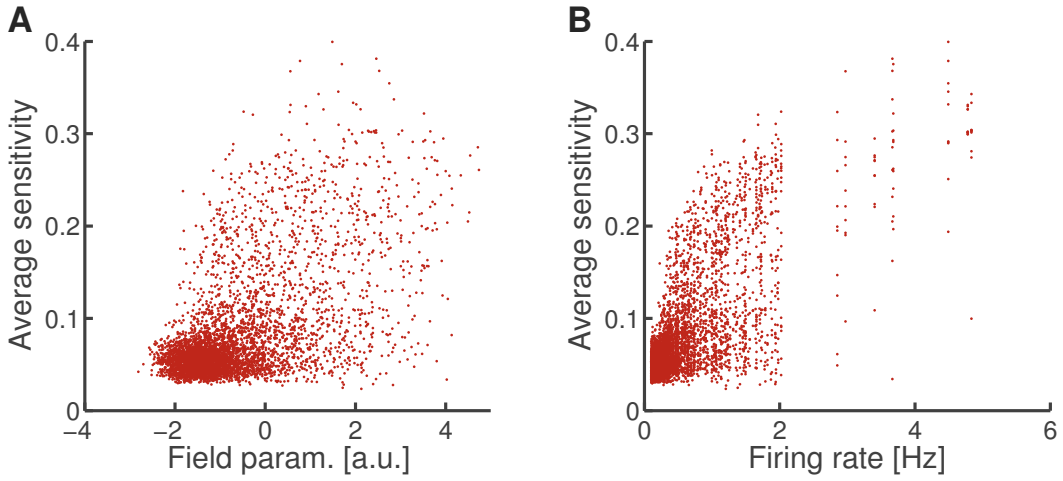

Figure 4.15: Relationship between average neuron sensitivity and its individual properties. (A) The averaged unit sensitivity scatter-plotted against the value of the field parameter, pooled across all groups in culture 1. (B) The averaged unit sensitivity scatter-plotted against the neurons firing rate, pooled across all groups in culture 1. Note the similarity with Figure 3.8A

Now, although the above results indicate that stiff dimensions tend to be dominated by a single or a pair of neurons, a question remains whether the situation is simply such that the neuron under consideration is an isolated unit, interacting only weakly with the group (rather than being a hub neuron as suggested earlier); and the reason for high sensitivity of its interactions is exactly due to its isolation, i.e. the groups' behaviour would be very different had the interaction parameters been higher. To

test for that scenario, yet another simple check was performed, similar in vein to the one pictured in Figure 4.14. Namely, the average strength of interactions of each of the units was calculated and then ranked according to the neuron's average sensitivity (see Figure 4.16A); and also a ratio of the average interaction strength of the most sensitive unit with respect to the average interaction strength of the rest of units was calculated (see Figure 4.16B).

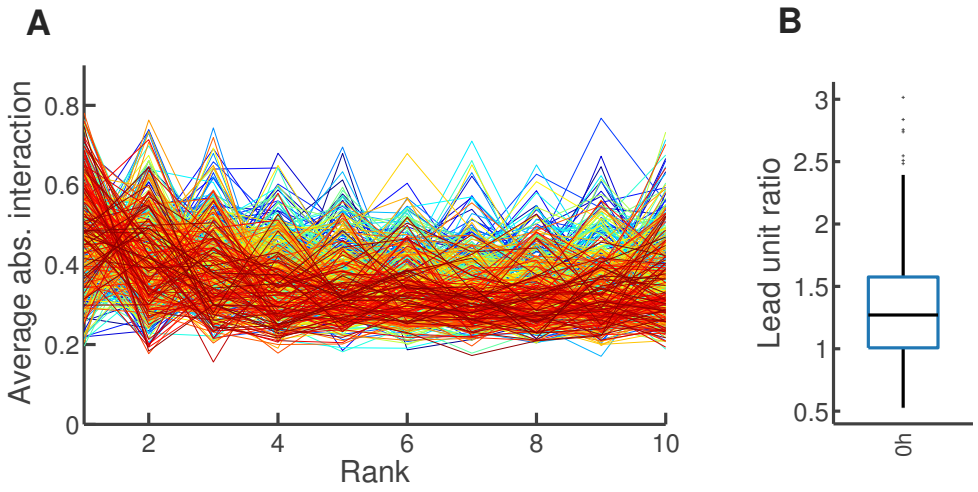

Figure 4.16: Average interaction strength per neuron in culture 1. (A) Averaged for each unit, the strength of interaction parameter (absolute value); units are ranked by magnitude of average sensitivity, as in previous figure; lines are color-coded according to the sparsity of the stiff dimension. (B) The distribution (across groups of neurons) of ratios of the average interaction strength of the most sensitive unit versus average interactions of remaining neurons.

Although the first panel of Figure 4.16 is somewhat difficult to read due to multiple lines crossing paths, this visual check suggests at least that the most sensitive neuron is not connected any weaker than the others. In some of the groups the average interaction of this first unit can be weak, but the overall trend seems to be in fact the opposite. For a more definite picture Figure 4.16B presents the boxplot of ratios of average interactions. Indeed, it appears that for about 75% of the groups the average interaction strength of the most sensitive unit is higher than the remaining neurons', although this difference is not very pronounced (on average it appears to be higher by about 25%).

This observation further solidifies the idea of 'hub' units: not only are they more active, but they appear more interactive too, making them indeed appear to constitute information hubs. It also potentially sets apart the present findings from such as those of Schneidman et al. (2006), where authors argue that strong dependency on the net-

work level can occur even at weak pairwise interactions. It is not disputed here that this can be the case, particularly in a different preparation (recall that their primary findings were from the salamander retina); however, in the present work the interaction parameters both reach higher values than those in Schneidman et al. (2006), and show a differentiation of interactions with unit sensitivity. Thus, it is suggested that different types of coordinated firing can arise from different underlying structures of networks; e.g. propagating retinal waves in more homogeneously connected retinal population, and highly globally coordinated bursts in dissociated cultures with hub neurons.

The results of sparsity analysis were consistent for all pairs of recordings and in each of the cultures, barring minor inter-preparation deviations. As can be observed in Figure 4.17, both cultures 2 and 3 exhibited the characteristic step-wise spectrum of sensitivities discussed above (shown in the figure are baseline recordings only from the 0h-2h comparison, but the same behaviour was observed across the data-sets). However, culture 2 tended to exhibit a more prominent presence of a third-ranked neuron contribution to the stiff dimension (see panel A of Figure 4.17), as well as a slightly smoother distinction between consecutive contributing units. Again, it is tempting to speculate whether this could have been brought about by a difference in cell clustering properties between preparations (recall that culture 2 was also the one exhibiting higher homogeneity and lower skewness of firing rates).

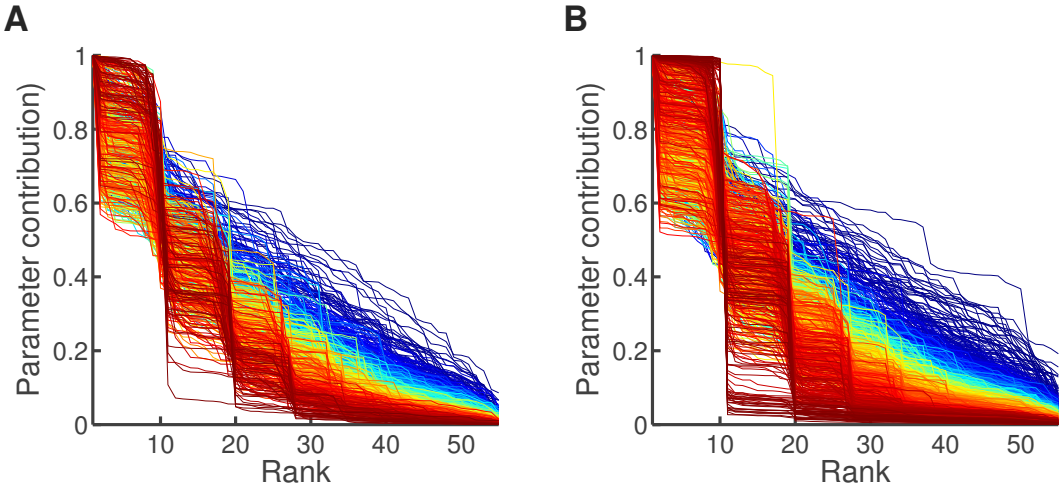

Figure 4.17: The distributions of individual parameter contributions towards the stiff dimension for the baseline recordings in cultures 2 and 3. (A) Ranked and scaled magnitudes of entries in the stiff dimension for all sets in culture 2; lines are color-coded according to the magnitude of their Gini coefficient. (B) Same for culture 3.

The distributions of Gini coefficients calculated for the sensitivities were also highly consistent across all recordings, as shown in Figure 4.18. As was the case for the baseline recording in culture 1 (see Figure 4.11B), indexes below 0.4 were uncommon, placing the majority of the groups above a random vector in terms of sparsity. The intermediate values of Gini index would suggest that the stiff dimensions tended to be composed primarily of two neurons, with a variety of balance distributions between them, depending on the group of neurons.

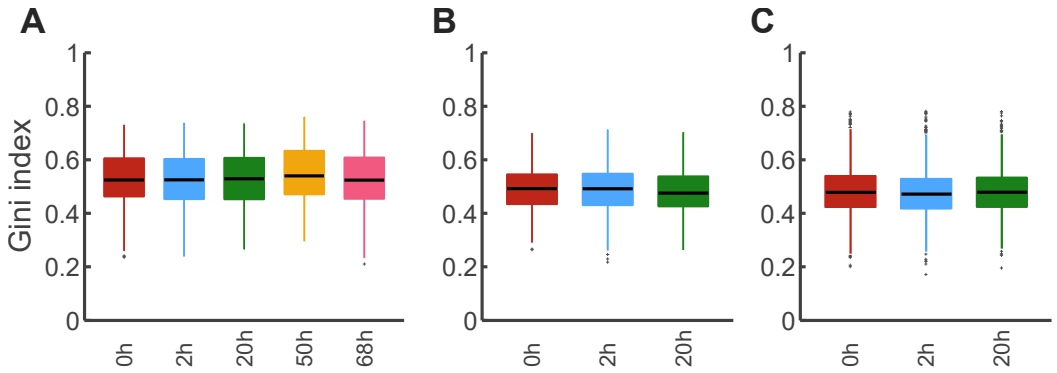

Figure 4.18: Sparsity of the models' stiff dimension across cultures and recordings. (A) Gini coefficients of the distributions of parameter sensitivities in culture 1; reported only baseline recordings from each pairwise comparison, as the Gini coefficients shown here are representative. (B-C) Same for cultures 2 and 3, respectively.

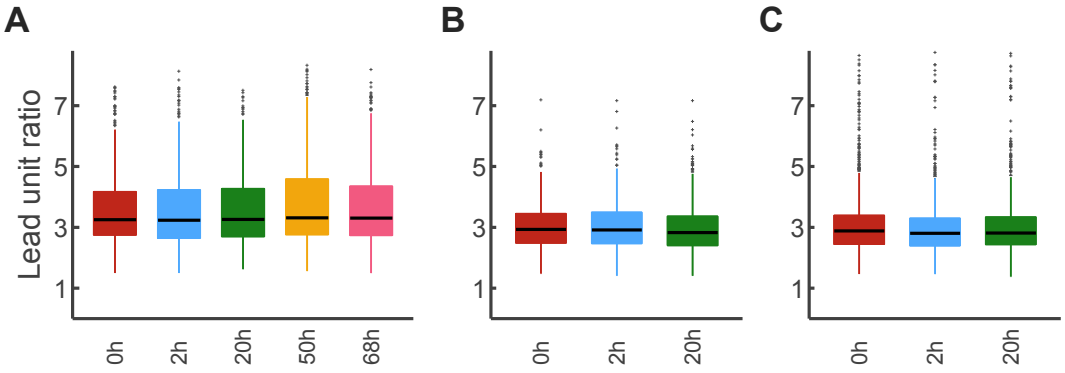

Figure 4.19: Distributions of the ratios of leading unit sensitivities to the average sensitivities of remaining neurons. (A) The average sensitivity across parameters of the leading unit divided by the average sensitivity across all other remaining neurons in a group, for all groups in the baseline recording from culture 1; reported only baseline recordings from each pairwise comparison, as the ratios reported here are representative. (B-C) Same for cultures 2 and 3, respectively.

As a further confirmation of the consistency of results, in Figure 4.19 reported for all cultures are the ratios of average sensitivities of the leading unit to the remaining

ones. In all cases for nearly all groups this ratio was higher than 2, i.e. the dominant neuron was at least twice as sensitive as the average over other units. For most cases in culture 1 this ratio was higher than 3; for cultures 2 and 3 this held for about half of the groups. The fact that only half of the sets exhibited ratios higher than 3, together with the picture in Figure 4.17, suggests that in those preparations there was a higher prevalence of mixed groups (i.e. with mixed components from two neurons, rather than a clearly dominant single one).

4.3.3 Comparing models across time

Since the pairwise maximum entropy models exhibited sloppiness across all preparations, recordings, and groups, it could then be examined whether it had any relation to the changes observed between time points. Specifically, it was of interest here whether there was any support for the hypothesis that differentiation of sensitivities is exploited by groups of neurons in maintaining stability. To provide first some illustrative clues, in Figure 4.20A the projections of parameters onto the eigenvectors from two consecutive recordings are pooled across groups and scatter-plotted against each other, color-coded by eigenvector rank (a 'flattened' version of Figure 4.6). It appears that the pattern observed in the example group shown in Figure 4.6, that is an increasing spread with increasing sloppiness, is consistent across examined sets of neurons. The stiffest projections (color-coded blue in the figure) all place along the diagonal, indicating very little changes took place along these directions. However, with growing sloppiness the scatter-plot widens considerably, and the projections no longer closely match between time-points.

When averaged across all groups of neurons, the relationship between eigenvector rank and magnitude of change can be clearly discerned (see panel B of Figure 4.20). Interestingly, it is not only the first eigenvector that is associated with little changes but the first several; at that point, however, a sharp increase in changes can be seen, followed by a slowly ascending trend until the sloppiest eigenvectors. The wide spread of projection changes indicated by the shaded region corresponds to standard deviation. While it might appear to stem from variability between groups, where some exhibit vast changes in the sloppy region, and some very small, the explanation is yet more involved. In fact, nearly all groups showed pronounced changes along at

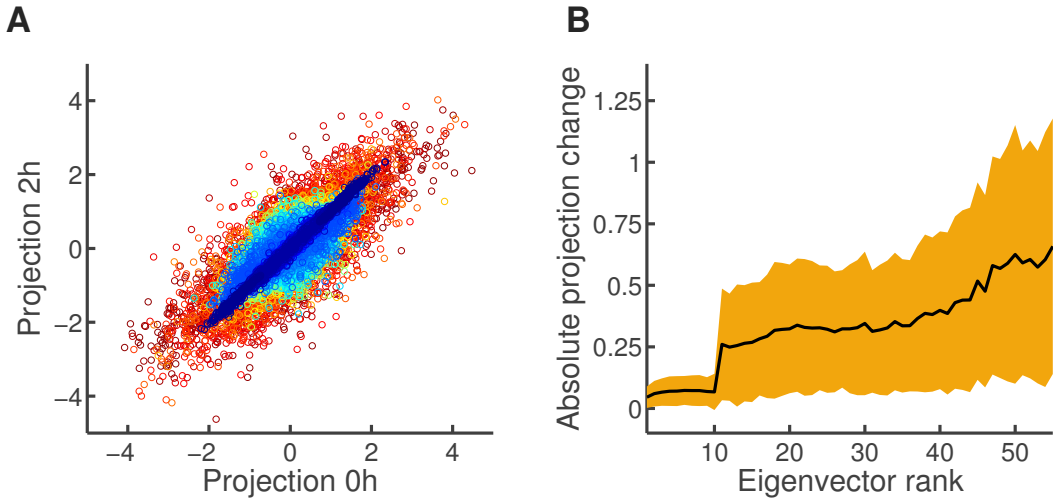

Figure 4.20: Comparison of model parameter projections between the first and second recording in culture 1. (A) A scatter-plot of parameter projections at 2h vs. 0h across all eigenvectors and all groups; color-coded according to the rank of eigenvector used for projection, as in Figure 4.6. (B) Average magnitude of projection change between the considered time points across all groups of neurons, plotted as a function of eigenvector rank; shaded region corresponds to standard deviation.

least some of their sloppy eigenvectors. However, among the approximately 40 insensitive eigenvectors, there were typically at least several that remained stable. Since the particular ranks for which this happened naturally differed from group to group, this gave rise to the high variability observed in Figure 4.20.

In order to qualify the above statement and further examine the robustness of the relationship between sloppiness and changes, for each group of neurons the magnitude of projection changes was computed separately for the sloppy and stiff subspaces (again employing the $Rk10$ criterion introduced in Section 4.2.4).

$$\Delta\Psi_{stiff} = \sum_{i=1}^{Rk10} |\Delta\psi_i| \tag{4.15}$$

$$\Delta\Psi_{sloppy} = \sum_{i=Rk10}^{55} |\Delta\psi_i|. \tag{4.16}$$

In this manner, it could be examined whether indeed in each of the groups there were significantly more projection changes occurring along insensitive, rather than the sensitive directions. As shown in Figure 4.21A, this was clearly the case here. Without exceptions, the amount of difference along sloppy directions vastly surpassed the changes along the stiff ones. Furthermore, this was not simply an effect of the size of

considered subspaces (recall that only few eigenvectors were necessary to account for 90% of the FIM sensitivity; in culture 1 it was typically 3-4 and never more than 8). As can be observed in panel B of Figure 4.21, the mean projection change per eigenvector was also significantly smaller for the sensitive subspace, and this held for every single group (see also Figure 4.22D). Therefore, the dominant amount of changes along the sloppy subspace was not simply due to a summation of many small contributions; in fact the individual contributions were larger than any of those occurring along sensitive directions. Thus, although the groups exhibited a wide range of projection changes, without exception the overwhelming majority of those took place in the sloppy subspace; and it appears that the magnitude of permitted change along stiff directions was constrained.

Importantly, the amount of change observed in the insensitive directions appears to be the major contributor to the parametric differences observed between recordings. As shown in Figure 4.21 C and D, there was a tight correlation between parameter similarity and total projection change in the sloppy subspace (Pearson r between the r^2 coefficient and$\Delta\Psi_{sloppy}$: -0.92), while the correlation with total projection change in the stiff subspace was pronouncedly lower (Pearson r between the r^2 coefficient and $\Delta\Psi_{stiff}$: -0.41). This observation is a further indication that indeed sloppiness plays a role in stability maintenance. It transpires from present results that the fluctuations and changes in individual parameters tend to predominantly take place within the confines of the insensitive subspace, i.e. along the flat areas of the manifold. Such changes, even of considerable magnitude, do not however influence strongly the collective group behaviour, since the eigenvalues associated with them are very low (recall the picture from Figure 4.9).

These results held also for all other recording pairs and each of the examined preparations. In Figure 4.22A-C, the ratios of stiff to sloppy projection changes are reported for all pairwise comparisons between time points, for cultures 1-3 respectively. As can be observed in these plots, without exception the magnitudes of changes within the sloppy subspace outweighed the stiff ones more than tenfold. There were minor differences between recording pairs, where the proportion of changes along the insensitive directions was increased from 0.01 to approximately 0.02-0.03, for example in culture 1 for the 20-44h and 44-50h comparisons. Recall from Chapter 2 that the recording at 44h was the performed 2 hours after a media change, therefore it

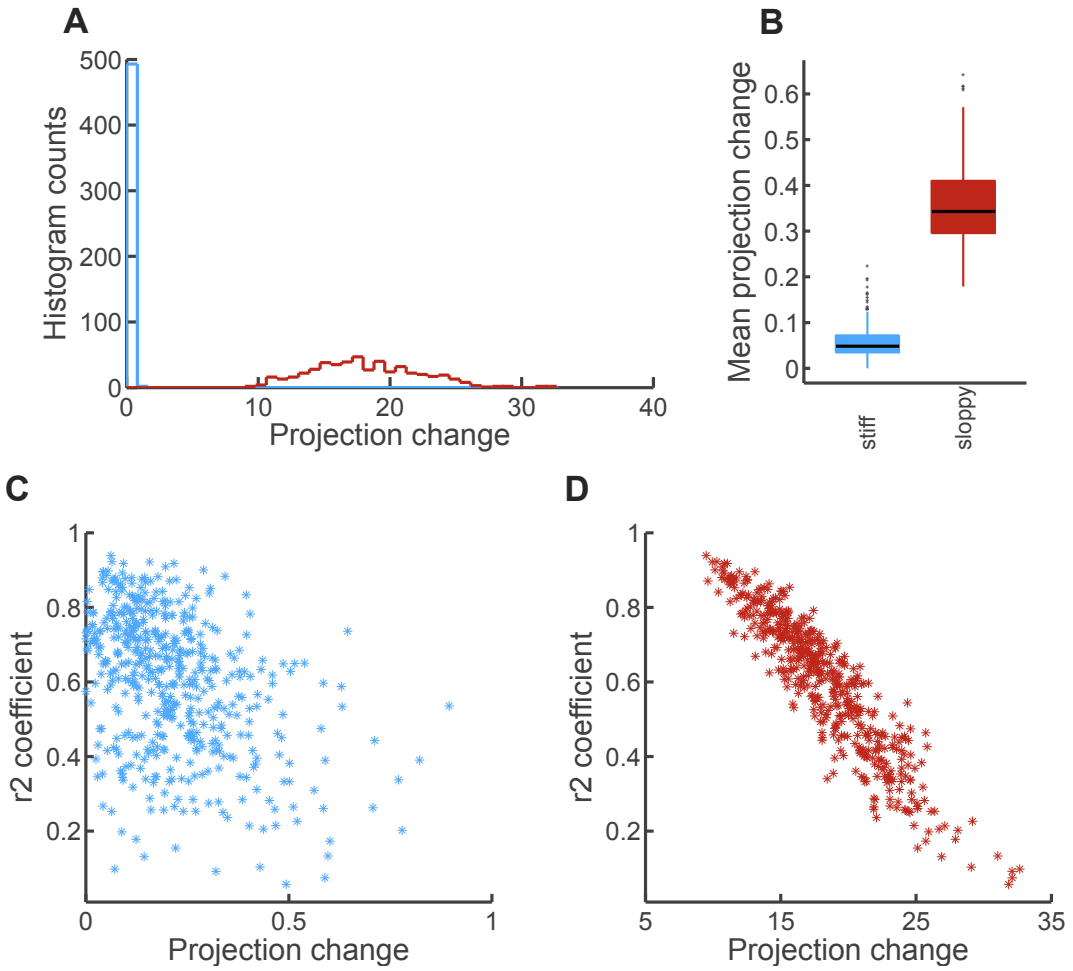

Figure 4.21: Comparing the distributions of projection changes per direction between stiff and sloppy subspaces. (A) Histograms across all groups of neurons in culture 1 of the total projection changes, separately calculated for the stiff and the sloppy subspaces as in Equation 4.16 (maroon and blue, respectively). (B) Boxplots of mean projection change per direction in the sloppy and stiff subspaces: the total magnitude of change per stiff / sloppy subspace was divided by the size of the subspace, to arrive at the average. (C) Relationship between parameter similarity (quantified by the r^2 coefficient of determination, as in the previous chapter) and the total amount of projection change in the stiff subspace, for each group of neurons from culture 1. (D) Relationship between parameter similarity and the total amount of projection change in the sloppy subspace, for each group of neurons from culture 1.

is likely that in that case some groups of units underwent more noticeable changes, including larger differences also along the sensitive directions. Nevertheless, overall the comparison of total changes falling into the stiff versus sloppy subspace, as well as the comparison of average changes per eigenvector (panels D-F of Figure 4.22 both indicate a robust effect of sensitivity).

Furthermore, also the relationship between the amount of projection change along insensitive dimensions and the parameter similarity between recordings was pre-

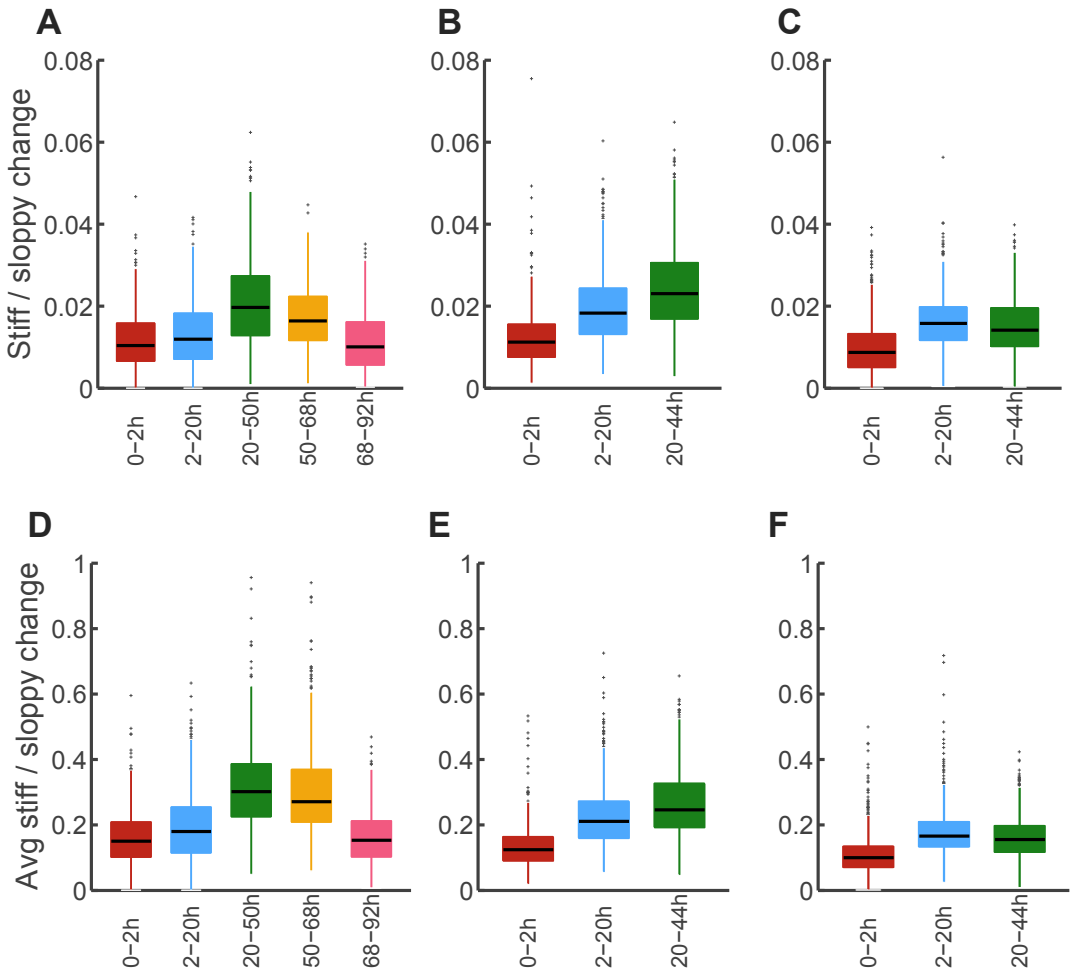

Figure 4.22: Distributions of ratios of projection changes along stiff and sloppy subspaces across recordings and preparations. (A) Boxplots of per-group ratio of total amount of projection changes in stiff versus the sloppy directions; consecutive pairs of recordings in culture 1. (B-C) Same as in panel A, for cultures 2 and 3, respectively. (D) Boxplots of per-group ratio of the average projection change per direction in stiff versus sloppy subspace; consecutive pairs of recordings in culture 1. (E-F) Same as in panel D, for cultures 2 and 3, respectively.

served across datasets. As reported in Table 4.1, the correlation coefficients for the above were close to 0.9. Thus, it becomes now clear how groups with even fairly dissimilar parameters could still retain highly similar distributions of spiking patterns.

A final question to address in the present work is whether the observed effects could have been brought about by the limitations of methods. Firstly, could the construction of the model implicitly result in a separation of sensitivities through two types of parameters; and secondly, how much of the sloppy changes were the result of fitting inaccuracies. In order to address these issues, two types of surrogate datasets were examined here: models with randomly permuted parameters and the

	0-2h	2-20h	20-44h	20-50h	50-68h	68-92h
culture 1	−0.92	−0.93		−0.91	−0.94	−0.94
culture 2	−0.90	−0.90	−0.87			
culture 3	−0.85	−0.86	−0.88			

Table 4.1: Correlation between the parameter similarity and the amount of projection changes between recordings occurring in the sloppy subspace.

re-sampled re-fit models re-purposed from the previous chapter (Section 4.2.5). In Figure 4.23 scatter-plots akin to that of Figure 4.20A are shown for illustration. Indeed, it can be observed in both cases that changes along the stiff directions (color-coded dark blue) are more restricted than those along the sloppy ones. Naturally, for the data-set with permuted indices these changes are far larger than either the original data or the re-sampled one. In turn, the re-fit re-sampled projections are overall noticeably more similar to the baseline recording than those of the 2h comparison in culture 1 (see Figure 4.24A), both in sloppy and stiff regions.

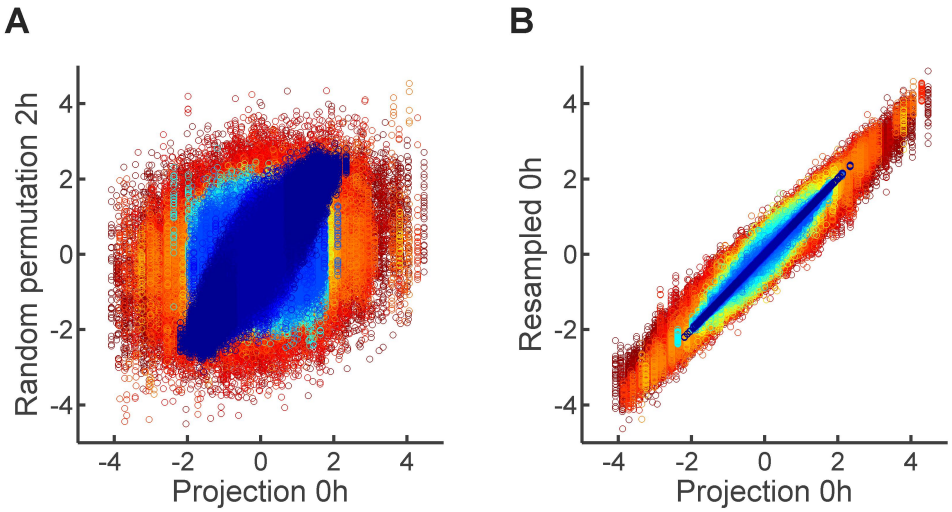

Figure 4.23: Comparison of model parameter projections to randomly permuted and re-sampled re-fit parameter projections. (A) A scatter-plot of parameter projections of randomly permuted parameters versus the original ones from baseline recording in culture 1; pooled across all groups and all 30 permutation runs; color-coded according to the rank of eigenvector used for projections. (B) A scatter-plot of projections of the re-sampled re-fit parameters versus the original ones from baseline recording in culture 1; pooled across all groups and all 30 re-sampling runs; color-coded according to the rank of eigenvector used for projections.

Thus, it appears that some degree of sensitivity separation is unavoidable in the type of model employed here. Therefore, in order to test to what extent the results of the present work were a true effect, the average changes per eigenvector in stiff and sloppy subspaces were re-computed, subtracting the averages estimated from the

re-sampled re-fit data. What transpires from this analysis is that although the fitting inaccuracies affected the sloppy subspace more than the stiff (as can be seen in Figure 4.23B), even after correcting for that imbalance the models fit to the data proved to change significantly more along the sloppy directions (Figure 4.24A). Again, similarly as for the uncorrected data presented in Figure 4.21, both the total magnitude of changes (Figure 4.24A) as well as average changes per individual direction (Figure 4.24B) were decidedly smaller for the sensitive subspace, indicating a constraint upon the malleability of the stiff dimension. This held true also for the remaining preparations, as presented in Figure 4.25.

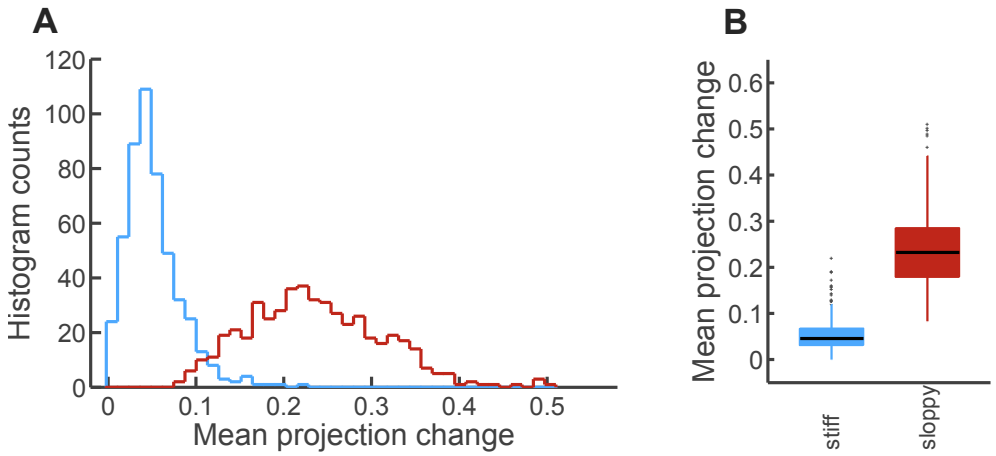

Figure 4.24: Comparing the distributions of projection changes in time per direction between stiff and sloppy subspaces, corrected for fitting inaccuracies. (A) Histograms across all groups of neurons in culture 1 of the magnitude of projection change per direction, separately calculated for the stiff and the sloppy subspaces (blue and maroon, respectively), after subtracting the corresponding values averaged across all 30 re-samples per set. (B) Boxplots of the corrected mean projection change per direction in the sloppy and stiff subspaces: the total magnitude of change per stiff / sloppy subspace, corrected by subtraction as in A, was divided by the size of the subspace, to arrive at the average.

Importantly, this effect was also significantly stronger than that for the randomly permuted data. As shown in Figure 4.25, the ratio of stiff to sloppy change per eigenvector was consistently lower than 0.5, save for outliers. In contrast, the same ratio was typically higher than 0.5 for the models with randomly permuted parameters. Thus, after correcting for fitting inaccuracies, the average insensitive parameter projection change was still more than twofold that of the sensitive, an effect that went beyond that of sensitivity separation between parameter types. In conclusion, it tran-

spires from the presented results not only that the models were sloppy and, to an extent, sparse, but that these properties were functionally relevant.

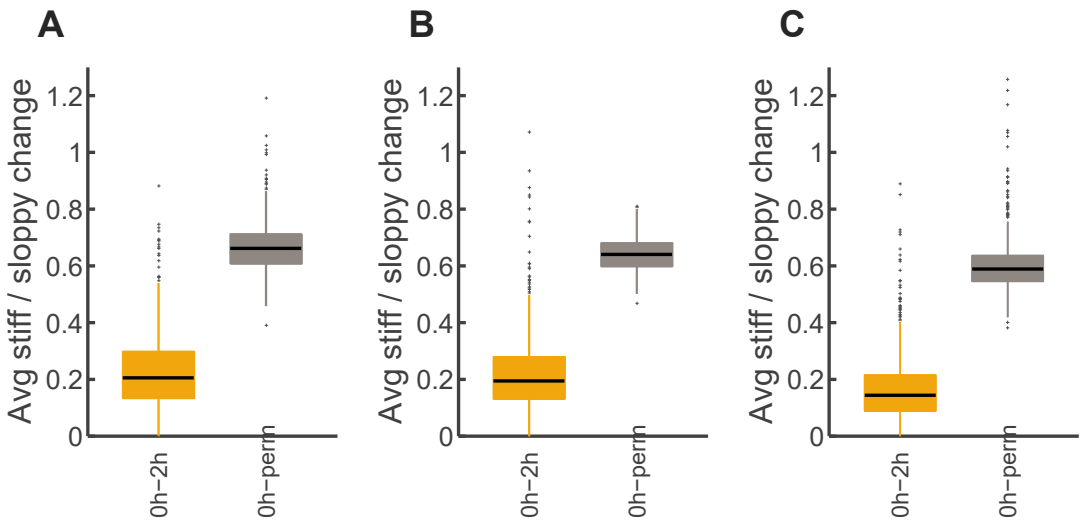

Figure 4.25: Distributions of ratios of stiff projection changes per eigenvector to the sloppy ones across all sets of neurons. (A) Boxplots of corrected () per-group ratio of average projection change per direction in stiff versus sloppy subspace (yellow) for the baseline recording in culture 1. Plotted in grey is boxplot of the same ratio pooled across the 30 permutations of parameters. (B-C) Same for cultures 2 and 3, respectively.

4.4 DISCUSSION

In this chapter of the thesis, information geometric methodology is employed to conduct a detailed analysis of the pairwise maximum entropy models introduced in Chapter 3. In particular, the Fisher Information Matrix is introduced as a tool for dissecting the sensitivity of probabilistic models to parametric changes (Amari and Nagaoka, 2007), and concepts from the area of sloppy model research are used to frame the results in intuitively understandable terms (Daniels et al., 2008; Transtrum et al., 2015). What transpires from this investigation is that the Ising models fit to recorded neural activity exhibit a property found in many other biological networks and physical systems termed sloppiness (Brown and Sethna, 2003; Daniels et al., 2008; Machta et al., 2013; Transtrum et al., 2015). That is, the collective behaviour of modelled neurons is governed by a small subset of *stiff* parameter combinations, and is insensitive to changes across a large *sloppy* subspace of parameter values. Furthermore, sparsity analysis of the stiff directions reveals that these combinations tend to be confined to a few neurons only. Importantly, sloppiness appears to be indeed a

functionally relevant feature for stability; that is, fluctuations in individual properties of neurons between recordings predominantly occur within the insensitive subspace, consistent with the proposed robustness hypothesis.

Employing factorization of the FIM to map the sensitivity profiles of the Ising models used in this work, it is shown here that these models exhibit signatures of sloppiness. However, since there exist no precise quantitative criteria to classify a model as sloppy or not, nor to measure the degree of sloppiness, direct comparisons to reported results are not possible. Instead, qualitative and operational similarities to other works are here examined; in addition, the Gini index is introduced as a concise measure of statistical dispersion, facilitating comparisons to reference distributions, and across examined models.

The eigenvalue spectra obtained across recordings and preparations exhibit the typical separation of sensitivities consistent with sloppiness, characterized by approximately regular spacing in the log domain, or geometric descent with rank (Brown and Sethna, 2003; Transtrum et al., 2015). While some individual groups show certain deviations from exact regularity (such as the example presented in Figure 4.5C, where two gaps separate the eigenvalues into three clusters), the key feature is always present: the majority of eigenvalues are orders of magnitude lower than the leading few. Upon visual comparison with reported distributions, the eigenvalues reported here typically decrease less steeply than e.g. for the diffusion model analysed by Machta et al. (2013), but in a comparable manner to systems biology models (Brown and Sethna, 2003; Daniels et al., 2008; Machta et al., 2013). Finally, a comparison against reference distributions reveals that for nearly all considered cases the statistical dispersion of eigenvalues is higher than that of a power law with exponent of -1.5, further confirming the assertion of sloppiness.

It is worth noting that the sloppiness observed in the pairwise maximum entropy models here is not an automatic property of the model class, nor are present results equivalent to those of Machta et al. (2013). The sloppiness demonstrated by these authors in the behaviour of the Ising model arises as the scale of observation is coarsened spatially; this, the authors argue, provides an alternative to the renormalization group approach (Machta et al., 2013). However, without the coarse-graining the original model is not sloppy (Machta et al., 2013). Here, in contrast, the models *are* sloppy and no coarse-graining was performed. The fact that in the present results a sepa-

ration of sensitivities is observed is, rather, due to the particular location upon the multi-dimensional manifold, imposed by the fits to the data.

Beyond establishing sloppiness, what was also of interest here was the particular nature of the stiff directions, as this has important implications for how stability can be maintained in groups of neurons. For example, in the case of sloppiness arising from coarse-graining examined by Machta et al. (2013), the emerging sensitive directions in parameter space are not just a subset of the original micro-parameters of the system. This makes sense from a physical point of view, since as the temporal or spatial scale becomes larger, local fluctuations cease to matter, and none stand out from the rest. In contrast, in certain systems biology models the dimensionality reduction occurs in a different manner, as certain individual parameters or indeed pathways in the network transpire to be irrelevant for collective behaviour and function (Brown and Sethna, 2003; Transtrum et al., 2015). A similar picture also emerges in the present work; although there is wide variability between particular groups of neurons, in the vast number of cases sensitivity can be traced to a few neurons and their interactions. Typically, at least half of the parameters and several units in a given set can broadly vary, without significant change in collective behaviour. Crucially, it was never the case that all parameters uniformly contributed to the stiff dimension. Thus, homeostasis can potentially be ensured through control of only a portion of the total population, with remaining units allowed to undergo plastic changes.

Whether indeed homeostatic plasticity operates in this manner cannot conclusively be established here. However, the results of comparison of modelled groups across time are indeed consistent with this picture. Although groups of neurons exhibited a range of parametric differences, consistently the majority of those occurred within the confines of the sloppy subspace (as summarized for culture 1 in Figure 4.21). This result was controlled for potential confounding factors related to parameter indeterminacy and the construction of the model (Figure 4.22). Thus, in parallel to the observations of Daniels et al. (2008), it is proposed here that robustness is enabled, or at least largely facilitated, by sloppiness: a wide array of individual neuronal properties can be subject to ongoing plasticity without destabilizing collective functionality, as long as the stiff neurons remain stable.

5

EXPLORING SLOPPINESS AND SPARSITY OF *IN VIVO* SPONTANEOUS ACTIVITY

5.1 BACKGROUND

The results reported thus far uncover a possible organising principle for orchestrating the mechanisms of plasticity and homeostasis in populations of neurons; however, they are of limited practical utility due to the nature of the data employed. The experiments conducted for this work were performed in dissociated cultures, and it is widely known that *in vitro* preparations, although useful in understanding many aspects of neural activity (Marom and Shahaf, 2002), exhibit also fundamental differences with respect to the behaviour of populations of cells *in vivo* (Corner, 2008). Notably, cultured cells randomly self-organize in a very different manner in comparison to the brain: they develop on a 2D surface, rather than in all 3 dimensions; they lack the biochemical cues from other areas of the brain that drive the development of laminar structure, functional differentiation, etc; and receive no input or feedback from the rest of the body and the environment (Corner, 2008). As a result, *in vitro* preparations lack the intricate structure of *in vivo* populations, something that is expected to result also in different activity profiles and range of behaviours. Indeed, studies such as those by Hinard et al. (2012) or Colombi et al. (2016) reveal that the natural repertoire of spiking in dissociated cultures does not resemble typical awake activity recorded *in vivo*. In fact, the characteristic bursting patterns observed *in vitro* (as discussed in Section 2.1) are broadly resemblant of the slow wave activity observed during deep sleep (Hinard et al., 2012; Colombi et al., 2016). In the deepest stages of sleep, large areas of cortical cells simultaneously undergo fluctuations in membrane potential (in the range of $0.3 - 1.5$ Hz), alternating between a hyperpolarized state, where spiking in the network is practically non-existent; and periods of de-polarization, where relatively normal level of neural activity is present (Steriade et al., 1993b), and, particularly, bursts of synchronised oscillations such as spindles

(Steriade et al., 1993a) occur. This results in atypically synchronous and – on average – slow activity, as compared to other states of consciousness. Interestingly, the similarities between the sleeping brain and spontaneously active dissociated cultures reach beyond basic activity measures: it appears also that cellular gene expression patterns characteristic of sleep are akin between *in vitro* and *in vivo* conditions (Hinard et al., 2012). Thus, it would appear that spontaneously active dissociated cultures might be representative of the sleeping brain.

Therefore, although neurons grown on a multielectrode array might maintain stability via mechanisms related to sloppiness, this does not necessarily hold for the activity of the living cells in the brain. Firstly, the modulation of activity via different states of consciousness, attentional mechanisms and task demands might each result in a parallel modulation of activity, inter-cellular communication, and expression of plasticity and stability. Thus, it is possible that global stability is maintained via different homeostatic mechanisms in e.g. the sleeping versus the waking brain. Secondly, complex laminar, columnar and regional organization of the brain can in principle exert additional constraints on stability. For example, although evidence suggests that the majority of neurons remain inherently plastic even through adulthood (Trachtenberg et al., 2002; Carmena et al., 2005; Ziv et al., 2013), some regions might be highly conservative, with nearly each neuron remaining individually stable (Chestek et al., 2007; Ganguly and Carmena, 2009). Thus, the hypothesis of stability via sloppiness needs independent validation in realistic *in vivo* settings.

Such tests are very challenging experimentally, for a range of reasons. Firstly, current recording techniques are still deficient when it comes to detailed and long-term / repeated monitoring of neural activity necessary for such analysis as employed here: implanted microelectrode arrays cannot guarantee consistent signals from the same individual neurons over spans of days (Lütcke et al., 2013), and are biased towards highly active units (Shoham et al., 2006) (which the present analysis suggests might be the stable ones dominating the stiff dimensions); and the alternative, 2-photon calcium imaging, while reliable in terms of cell identity, suffers from poor temporal resolution (Lütcke et al., 2013). A second issue arising here is even more problematic: the very fact that activity in the brain can be rapidly modulated poses a problem for any method aiming to uncover underlying properties of neurons. One would require to separate out periods of differently modulated activity, in order not to confound the

analysis by including simultaneously two different modes of behaviour. This poses a serious difficulty, both for the methodology presented in this work, as well as other approaches, unless complementary measurements are conducted simultaneously: it is impossible to determine from activity alone whether the time-varying behaviour of some neurons is a signature of their default state, or a result of e.g. neuromodulatory influence.

With these caveats in mind, it was attempted here to establish whether *in vivo* populations of neurons show signs of sloppiness in their activity, since this is a prerequisite for any further investigations of stability. In the present chapter, spontaneous activity of neurons recorded from primary visual cortex of a monkey (Chu et al., 2014b) is analysed in the same manner as the dissociated cultures, and the results are discussed in the context of implications for further investigations.

5.2 MATERIALS AND METHODS

5.2.1 Recordings of spontaneous activity *in vivo*

The neural activity recording used in the present chapter (Chu et al., 2014a) was obtained from a public neuroscience data repository, Collaborative Research in Computational Neuroscience (crcns.org), and used here with permission of the authors. The dataset was originally collected for the purpose of investigating the relationship between correlations and tuning properties of primary visual cortex neurons, as reported in Chu et al. (2014b), and comprised both spontaneous and evoked *in vivo* recordings from lightly anaesthetised macaque monkeys (*Macaca cyclopis*). Only the spontaneous activity data was employed here, since the evoked protocol introduces complicating factors with respect to current analysis (e.g. repeated presentation of stimuli constrain the repertoire of firing patterns). Full details of all experimental procedures are available in the original study (Chu et al., 2014b) (the experiments were performed in accordance with appropriate ethical guidelines and requirements); here a summary of key details of the materials and methods is provided.

The monkey was anaesthetised and a 64-site multielectrode array (8 shanks with 8 electrode contacts each; inter-electrode separation of 200μm; Neuronexus Technologies, Inc.) was inserted normal to the cortical surface in the primary visual cortex

(V1). The correct positioning of the array within V1 was assessed independently via imaging and electro-physiological measures. Spontaneous neural activity was then recorded under light anaesthesia for 15 minutes, filtered ($400 - 5000$Hz, 48dB per octave) and digitized at 24.4kHz. Subsequently, spike sorting was performed off-line to separate out the activities of individual neurons, employing the super-paramagnetic clustering algorithm (Quiroga et al., 2004) with manual adjustments of sorting parameters. After additional manual corrections (such as excluding clusters with spike intervals violating the refractory period), 140 separate active cells were identified.

5.2.2 Data preparation and analysis

In order to adapt the above recordings to a spike file format compatible with the algorithms used in the thesis and to allow for a similar visualisation as for the *in vitro* data, electrodes from the *in vivo* study were assigned a location on the 64x64 MEA layout. The inter-electrode spacing of the Neuronexus array was 200μm, which roughly corresponds to a distance of 5 electrodes on the CMOS MEA; therefore, electrodes from the *in vivo* study were assigned positions from electrode [10,10] in steps of 5 up to [45,45]. However, recall that spike sorting was performed by the authors, since multiple neurons could be recorded from a single electrode. As there was no positioning data available regarding these neurons with the respect to the recording electrode, the positions of individual cells were assigned arbitrarily within the neighbourhood of the central electrode [x,y], with the first mapped to [x,y], second to [x,y+1], etc., i.e. forming a cross around the central electrode (since there were no cases of more than 5 units identified on any single electrode).

Once appropriately formatted spike files were obtained, the rest of the analysis, i.e. pre-processing and binning, Ising model fitting and Fisher Information Matrix decomposition, were performed exactly as described in previous chapters, for a 100 randomly chosen sub-groups of 10 neurons. The re-sampling re-fitting procedure was here performed 5 times per each set of units, as compared to 30 instances of re-sampling for the dissociated cultures, due to time constraints.

5.3 RESULTS

The results reported in this chapter were also used in the publication closely related to the thesis (Panas et al., 2015); however, the results reported here are more detailed and the performed analysis was extended with respect to the published version (e.g. by investigating stationarity) and updated with a newer version of sloppiness analysis (i.e. the construction of the stiff matrix here included more than just the first eigenvector, while in the Panas et al. (2015) paper only the first eigenvector was employed as a proxy for sensitivity).

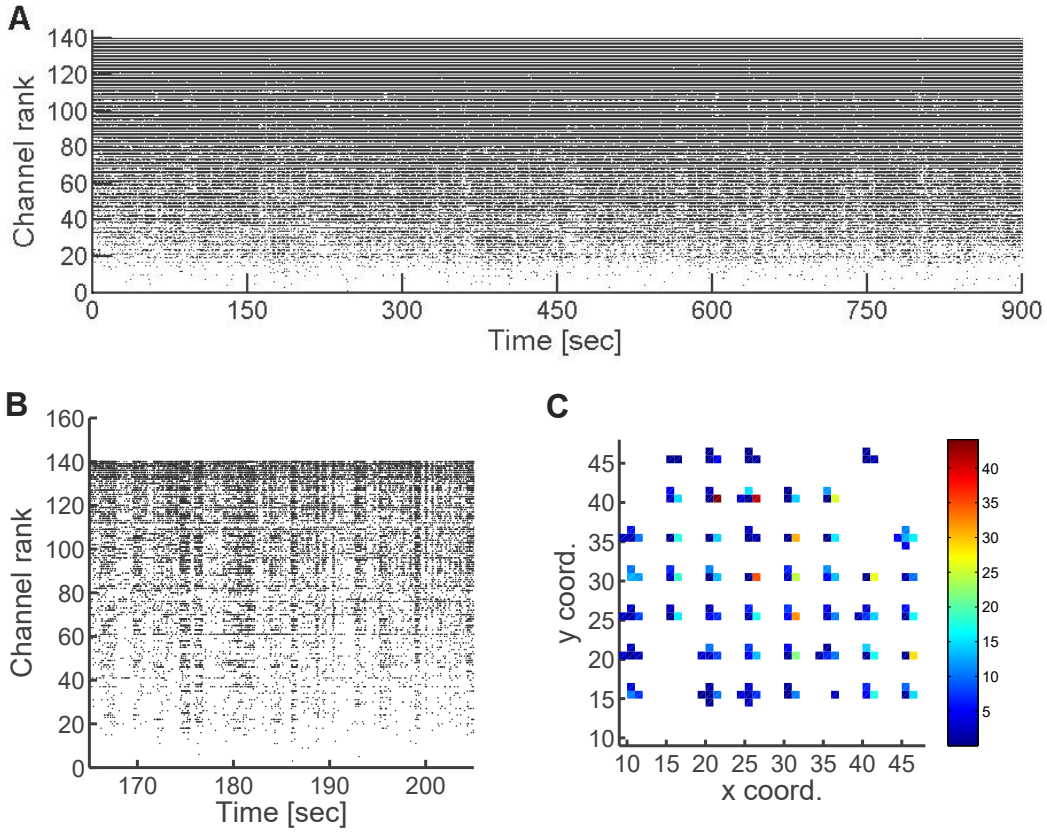

Figure 5.1: Illustration of the recorded spontaneous *in vivo* activity of neurons in the primary visual cortex of the macaque monkey. (A) Raster plot of individual spikes in each of the 140 identified neurons, with neurons ranked by their firing rate; each dot denotes a spike. (B) A close-up of 40 seconds of activity in the recording, highlighting the synchronous aspect of neural activity not apparent in the full raster shown in A; however, note the difference with respect to culture activity (see Figure 2.10), in particular much higher and denser activity and lower degree of coordination amongst units. (C) Approximate layout of the recorded neurons (see Section 5.2.2 for details of assigning neurons position); color code denotes the firing rate of a given neuron [Hz].

5.3.1 General properties of spontaneous neural activity *in vivo*

First, in Figure 5.1 an illustration of the spontaneous *in vivo* activity of the population of neurons in monkey V1 is presented. A visual examination of the raster plot already confirms the points raised in the introduction: namely, the fact that the behaviour of *in vitro* cultures is not representative of the ongoing spontaneous activity in the intact brain. A comparison with Figure 2.9 reveals two major differences: activity of cells in monkey V1 is much more asynchronous than the cultures'; and the firing rates of neurons in the primary visual cortex are much higher and span a decidedly wider range than those of *in vitro* activity. A closer look at the raster, shown in Figure 5.1B, provides hints as to the source of that difference: while there are still bursts of coordinated activity present in the *in vivo* recording, they are interspersed with ongoing de-synchronised activity (as opposed to silence that occurs in the cultures, see Figure 2.9).

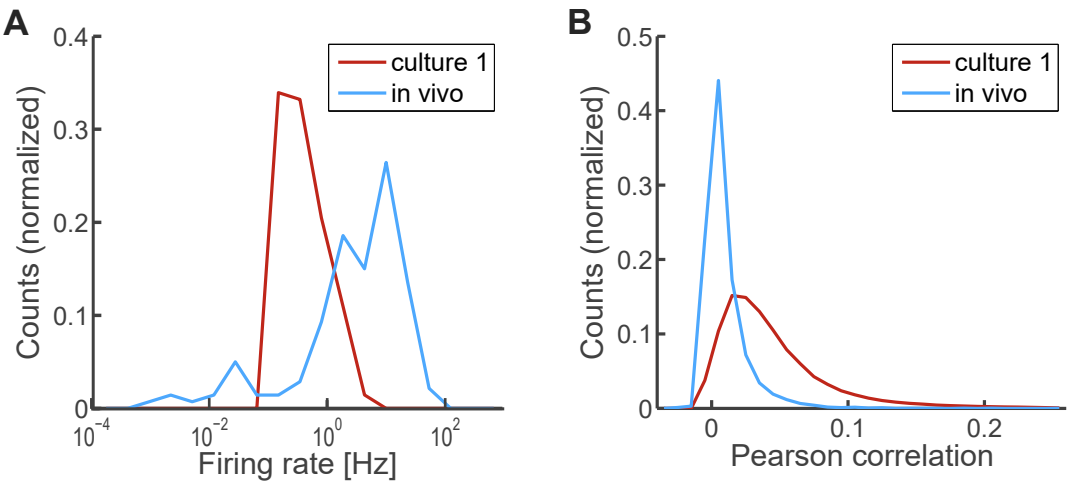

Figure 5.2: Distributions of firing rates and correlations in the recorded spontaneous *in vivo* activity, compared to that of an example *in vitro* recording. (A) Histogram bin counts of firing rates across neurons; plotted on a log scale for ease of visual assessment. (B) Histogram bin counts of Pearson correlation coefficients between all pairs of neurons (Pearson correlation coefficients calculated for spikes binned into 5ms time bins).

A direct comparison of the distributions of firing rates and correlations between the *in vivo* and representative *in vitro* activity (first recording from culture 1), shown in Figure 5.2, agrees with the above observations. As can be seen in panel A, the firing rates in the culture are more constrained and span a much narrower range; crucially, none of the recorded *in vitro* neurons exhibit spiking higher than 10Hz,

while in the recording from monkey V1 it is the majority of cells that fire above that threshold. On the other hand, *in vivo* both the range of observed correlations is much lower than *in vitro*, and the average and median correlations are clearly lower as well. Taken together, these plots confirm the intuition that while there is more activity in general in the recording from monkey V1, a much smaller proportion of that activity is coordinated.

5.3.2 Evaluation of maximum entropy models, and links to asynchrony and nonstationarity

The differences observed between the spontaneous activity obtained from the culture and that from the monkey V1 already suggest that likely the behaviour of the model fits will be somewhat different between the two types of data. For instance, it is expected that fewer all-encompassing bursts will result in a negligible level of higher-order correlations, leading to a better Ising model fit; and that on the other hand the apparent asynchrony will translate to better fits of the independent model, and perhaps a lower advantage of the Ising model over the independent one. As shown in Figure 5.3, these intuitions do appear to hold at first glance.

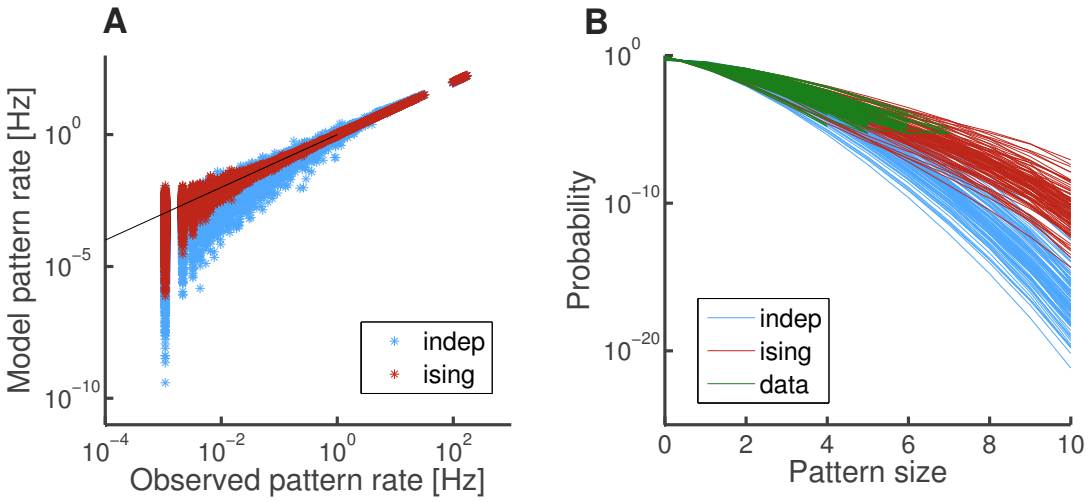

Figure 5.3: Illustration of the predictive power of the maximum entropy model. (A) The modelled probabilities of each possible firing pattern in a set of 10 neurons are plotted against the observed probabilities; data pooled from all sets of neurons in the analysed *in vivo* recording; for reference pictured in blue are the probabilities predicted by the independent model. (B) The distributions of the predicted and observed sizes of spiking patterns.

First, note that the predicted pattern occurrence matches the data better than was the case for the cultured data, for both of the employed models (compare panel A with that of Figure 3.4). Second, although still on average the Ising model offers an improvement over the independent one (see panel A), the Zipf plots shown in panel B exhibit signs of the independent model closely resembling the data in some of the examined groups. Compare these plots to those shown in Figure 3.4, for a clear demonstration of how the independent model is much more deficient when it comes to *in vitro* data (for example, note that in panel B there is a clear separation of the predicted probability of pattern size 4, which is not present in Figure 5.3B).

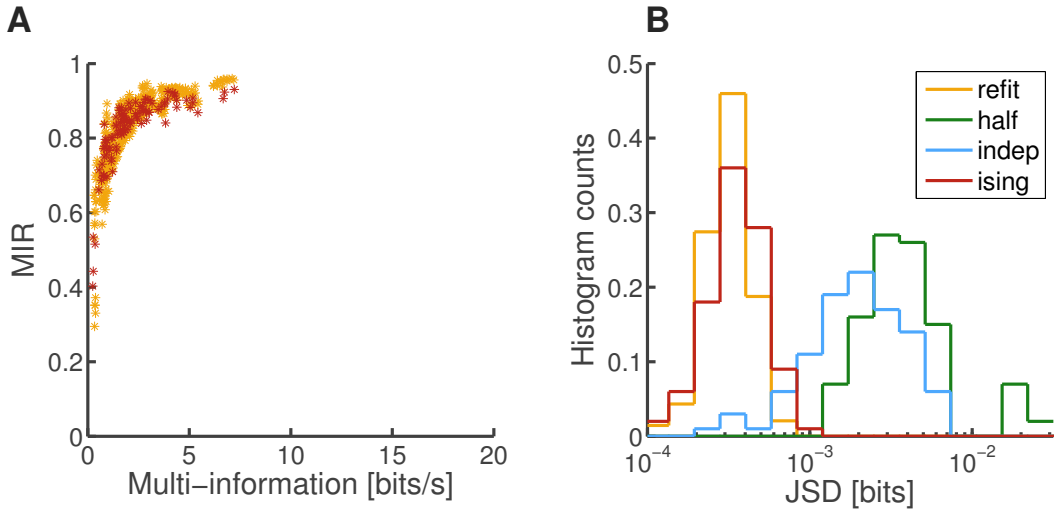

Figure 5.4: The distributions of measures of model fit quality in the recording from monkey V1. (A) The model multi-information ratio plotted as a function of multi-information for each set of neurons; also plotted for reference the multi-information ratio of re-sampled re-fit dataset. (B) Normalized histogram counts of Jensen-Shannon divergences of independent models for same groups of neurons, as well as divergences of distributions derived from halves of the data and those of the re-sampled re-fit dataset.

However, these are more of auxiliary visualisations; and in order to examine this directly, in Figure 5.4 plotted are the main measures of fit quality: the multi-information ratio for the Ising vs. independent model (as a function of multi-information), panel A; and Jensen-Shannon divergences between modelled and observed distributions (as well as reference ones), panel B. Although indeed for several groups with low multi-information the Ising model does not outperform the independent one as prominently as for the rest of the groups (with MIR being as low as 0.4), or the dissociated cultures (see panel A of Figure 3.5), there are few such cases only. These rare groups are presumed to consist of largely uncorrelated units, well described by the indepen-

dent model (hence little gain from using the Ising model); the low multi-information on the x-axis itself is an indicator of this fact, since low multi-information means that the entropy of the independent model closely matches that of observed data distribution. The rest of the groups, however, show very similar MIR to *in vitro* data, indicating that despite its asynchronous appearance, the *in vivo* recordings are not in fact well described as independently firing units. What is worth noting here is that the Ising model performs practically the same as the re-sampled re-fit models (which are intended to provide an approximation as to how well models can match data, given that there are limited samples in the recording). This means that, unlike in the dissociated data, here the maximum entropy model was able to capture practically all the compound correlation in the groups.

For further confirmation, note the distributions of Jensen-Shannon divergences, presented in panel B of Figure 5.4. Here it is again apparent that the Ising model outperforms the independent one in most groups; and even though the range of multi-information in the *in vivo* data is lower than that of *in vitro*, the distributions of JSD for the independent models are quite similar between the two types of data (compare with panel B of Figure 3.5). The Ising model, on the other hand, tends to fare somewhat better here than *in vitro*, both in terms of bits, but also in reference to the re-sampled re-fit JSDs (which provide a lower bound of how well the distributions can be predicted under the given sampling conditions). This is again conforming with the intuition that higher-order effects play even less of a role in this type of data and the *in vivo* recordings are noticeably less synchronous overall. Altogether, it seems so far that the maximum entropy model provides a rather good description of the monkey V1 recording.

However, a troubling issue arises here when examining the distribution of Jensen-Shannon divergences between halves of the recorded activity (denoted in green in Figure 5.4B). Not only does the Ising model outperform the data-data prediction, but even the independent model tends to do so in many of the cases. Also, comparing the data-data prediction here with that of dissociated cultures (see panel B of Figure 3.5) makes it apparent that this is not the case of slight overfitting, which might be expected due to half-data having fewer time bins as basis of probability distribution estimates (especially that given generally higher firing rates *in vivo* the sampling bias is expected to be less pronounced, rather than more). To better understand this issue,

it is instructive to employ an additional comparison; it is also a data-data prediction, but based in this instance not on halves of the recording, but on bins that are randomly permuted and then split into two datasets. In Figure 5.5 such comparisons are presented both for the dissociated culture data and the monkey V1 recording.

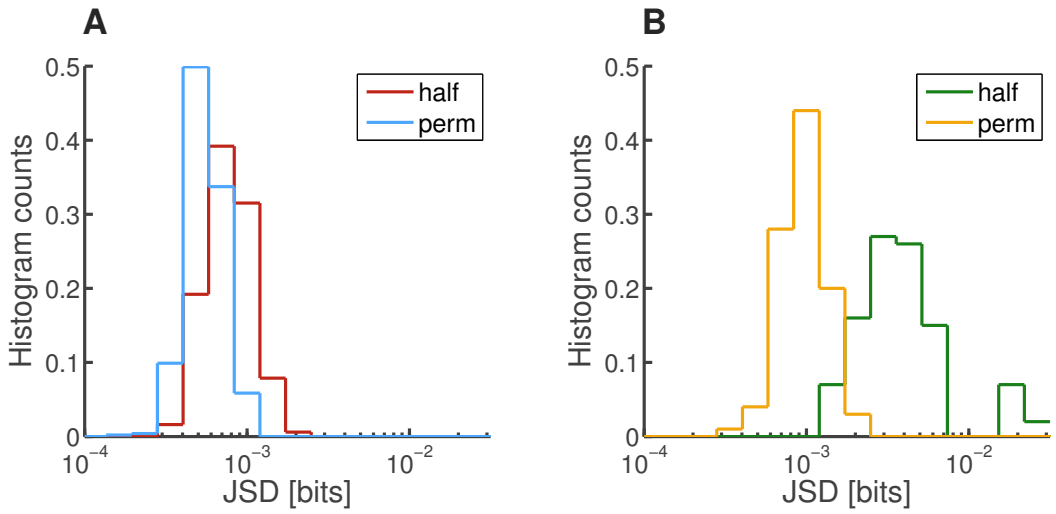

Figure 5.5: Investigation of the predictive power of halves of the data. (A) Normalized histogram counts of Jensen-Shannon divergences between spiking pattern distributions predicted from either first vs second half of the data (denoted 'half'), or random half bins vs remaining bins (denoted 'perm'); data for the *in vitro* recording culture 1. (B) Normalized histogram counts of Jensen-Shannon divergences between spiking pattern distributions predicted from either first vs second half of the data (denoted 'half'), or random half bins vs remaining bins (denoted 'perm'); data for the *in vivo* recording from monkey V1.

The function of employing randomly permuted bins versus half-half comparison is to establish a benchmark of deviations that can be expected from the fact that only half of the bins are employed for inferring probability distributions (thus lowering the number of samples). As can be seen in panel A of Figure 5.5, for the *in vitro* recording these two distributions of divergences largely overlap, meaning that whatever deviations occur between the first and second half of the recording are largely random sampling fluctuations. Not so, however, in the case of the *in vivo* recording – something that might be found somewhat counterintuitive, because as the reader might recall, the raster plots of dissociated cultures did at times look nonstationary, whereas the asynchronous raster of monkey data appears to the eye to be fairly self-similar throughout. Here though, the randomly permuted halves clearly and by an order of magnitude outperform the half-half distribution matching. This indicates, essentially, nonstationarity: the networks visit different states with different rates in the first versus the second part of the recording.

This is exactly an example of the issue flagged in the Section 5.1 – the confounding effect of rapid modulation of activity in the brain. The change in the distributions of firing patterns between the first and second halves of the *in vivo* recording could have arisen from a range of possible causes; and a key distinction here for drawing any conclusions is whether it was a discrete change in the state of the population (e.g. change in attentional state) or rather an expression of a natural time-course of evolving patterns. If the former were to be the underlying cause of nonstationarity, any further analysis is essentially futile, as the model fits would be performed on an aggregation of data from two distinct modes of operation. However, should it be the latter case, such nonstationarity does not necessarily present an issue – although it still requires careful follow-up investigations to definitely claim that. Essentially, spike patterns in neural populations can be analysed in terms of statistics of their spatial arrangement; an oft used analogy in this context, as already mentioned, is to treat them as a collection of population code-words, each representing something in the neural dictionary (and each with different rates of occurrence, as is the case with any words of a dictionary). Just as words in a sentence, these patterns might have a tendency to follow one another with a distinct temporal statistic, which might result in certain apparent nonstationarity; this however does not preclude studying the words themselves, as long as it is ensured that the observation time is sufficient to construct a robust estimate of the distribution of those spiking patterns.

For the purpose of further analysis, it is assumed here that the nonstationarity observed in the *in vivo* recording was an instance of a time evolution of spiking patterns, rather than a distinct change in the brain state (given that the monkey was maintained under light anaesthesia, it is unlikely that a major shift in attentional / consciousness state occurred during the recording). However, this is purely an assumption and one impossible to verify without further experiments. Therefore the results reported herein should be treated as a proof of concept / lack of contradictory evidence, rather than definite support for the posed hypothesis.

5.3.3 Signatures of sloppiness and sparsity in spontaneous cortical activity

Having established adequate quality of Ising model fits, it was then possible to proceed with calculations of the Fisher Information Matrices, and subsequent analyses

of sloppiness and sparsity. First, shown in Figure 5.6 are the distributions of eigenvalues resulting from FIM decompositions for each of the model fits. This picture closely resembles the one resulting from analysis of the culture data (compare with Figure 4.7), indicating that also *in vivo* recordings exhibited signatures of sloppiness. The eigenvalue distributions for the most part show the signature steep decline on a log scale, and Gini coefficients average close to 0.87, although there were a few outlier groups.

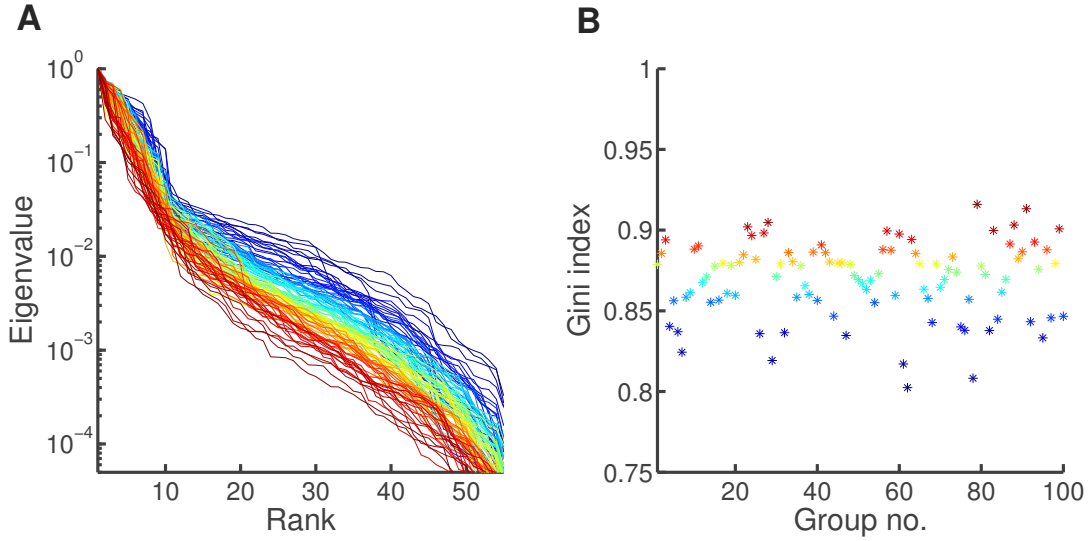

Figure 5.6: The distributions of eigenvalues and corresponding Gini coefficients from the *in vivo* recording. (A) Eigenvalue plotted as a function of its rank for each model fit; eigenvalues are scaled to the first one, for ease of comparison; lines are color-coded according to the magnitude of their Gini coefficient, see panel B. (B) Gini coefficients calculated for each of the distributions from panel A.

For a more direct comparison, in Figure 5.7 the distributions of Gini coefficients obtained from the monkey V1 data are compared to those of a representative dissociated culture, as well as the two simulated data-sets of independent Poisson units (see Section 4.2.4). As can be seen from these boxplots, *in vivo* data differs from *in vitro* only marginally, while both are statistically different from the simulated and modelled independent surrogate sets (as discussed in Section 4.3.1). Such a picture suggests that groups of neurons in the cortex exhibit the necessary prerequisite of a sloppy system, i.e. a differentiation in parameter sensitivities.

In order to determine whether this sloppiness was consistent with the hypothesis of distinct sensitive units, the second step in the analysis was to examine the sparsity and structure of the stiff dimension. In Figure 5.8A, individual parameter

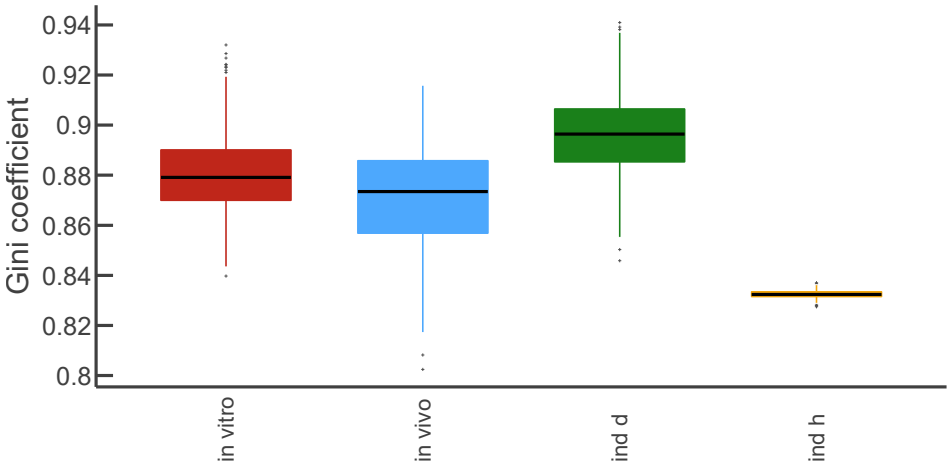

Figure 5.7: Comparison of sloppiness across preparations and simulated data. Boxplots of Gini coefficients of eigenvalue distributions for each set of modelled neurons in: culture 1 (denoted 'in vitro'), monkey V1 (denoted 'in vivo'), data generated from independent models matching the firing rates of culture 1 (denoted 'ind d'), and data generated from independent models with all equal firing rates (denoted 'ind h').

contributions towards the stiff matrix (ranked and scaled) are shown, analogously to Figure 4.11 and Figure 4.17, exhibiting again a very similar pattern to the one exhibited by models from dissociated cultures. In particular, the typical step-like curves are present, suggestive of dominance of 'hub' units over group behaviour.

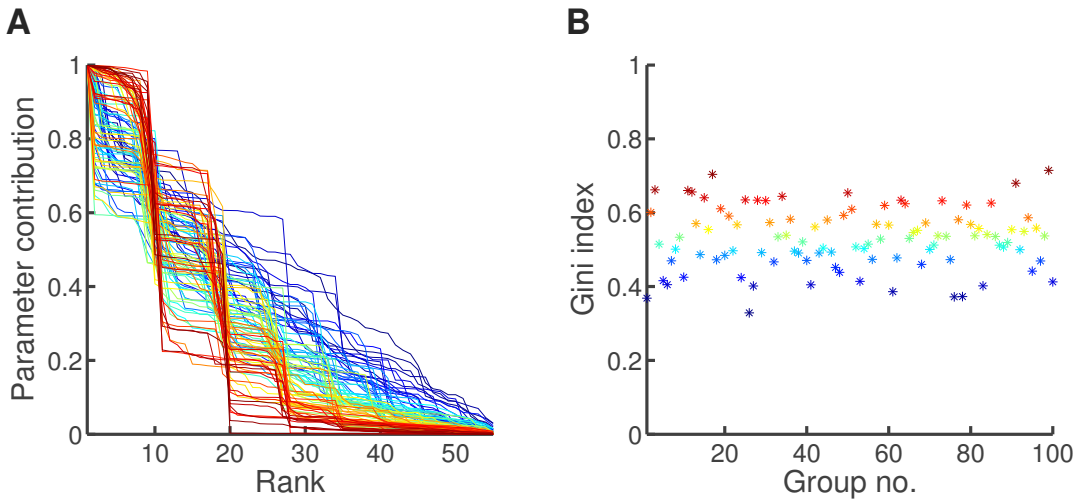

Figure 5.8: The distributions of individual parameters contributions towards the stiff dimension and the corresponding Gini coefficients for the models fitted to monkey V1 data. (A) Ranked and scaled magnitudes of entries in the stiff dimension (a weighted sum of the leading eigenvectors), i.e. essentially a distribution of sensitivities of individual parameters for each set of neurons; lines are color-coded according to the magnitude of their Gini coefficient, see panel B. (B) Gini coefficients calculated for each of the distributions from panel A.

This can be further confirmed by examining the average sensitivity per neuron in each group, as shown in Figure 5.9. As evident from this plot, the sensitive parameters indeed cluster to individual neurons, outranking the remaining units by a ratio of approximately 3 on average; this essentially replicates the results demonstrated for the dissociated cultures in Figure 4.14. Also the Gini indices calculated for the stiff dimension appear very similar to those of the *in vitro* (see panel B of Figure 5.8); however, for a direct comparison let us turn again to boxplots comparing *in vitro*, *in vivo* and simulated Poisson data.

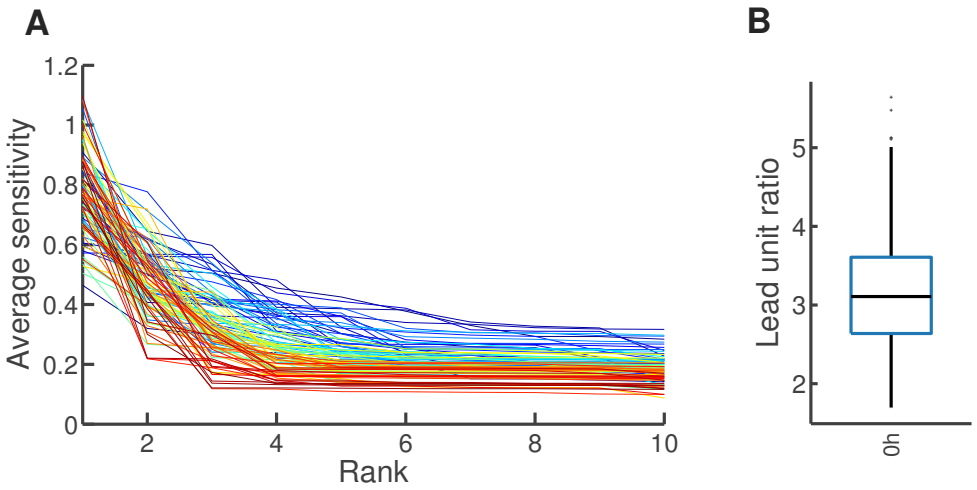

Figure 5.9: Average parameter sensitivity per neuron in monkey V1. (A) Averaged for each unit, contributions of individual parameters towards the stiff dimension, ranked by magnitude; lines are color-coded according to the sparsity of the stiff dimension. (B) The distribution (across groups of neurons) of ratios of the average sensitivity of the leading unit versus average sensitivities of remaining neurons.

As can be seen in Figure 5.10, indeed the sparsity of the stiff dimension is closely resemblant between the two types of recorded data. In turn, the Gini indices of the models fitted to simulated homogeneous Poisson spikes are, predictably, significantly lower than the rest; this is due to the fact that there is no reason for any single unit to be more influential than the rest (confirming what was discussed in Section 4.3.1). However, sparsity measure distributions of the models fitted to simulated Poisson spikes with data-matched firing rates are nearly indistinguishable from the recorded data.

Again, as discussed earlier, there is a simple explanation for that fact: both the recorded data, and rate-matched Poisson data, feature units of varying firing rates. The highly active units are expected to be more influential than the others, and that

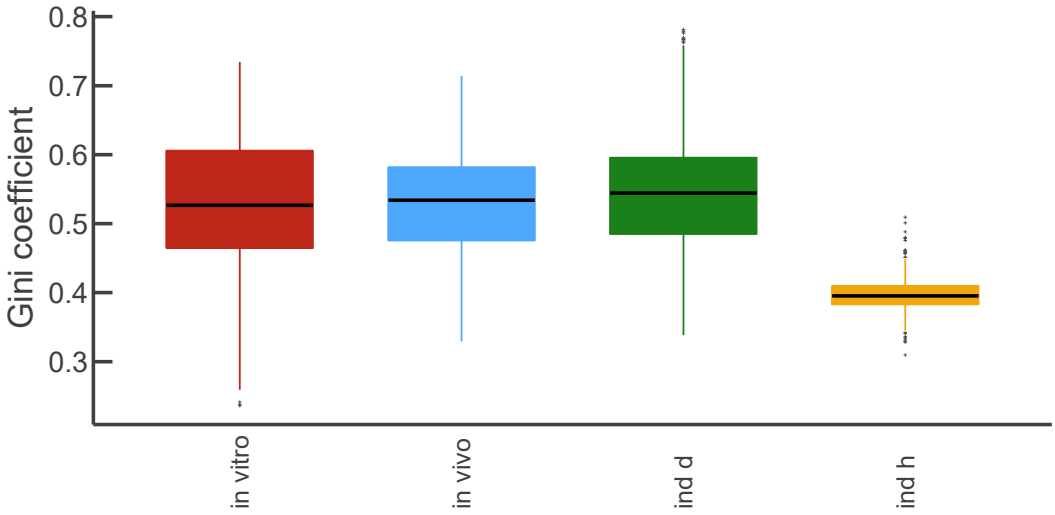

Figure 5.10: Comparison of sparsity of the stiff dimension across preparations and simulated data. Boxplots of Gini coefficients of parameter sensitivity distributions for each set of modelled neurons in: culture 1 (denoted 'in vitro'), monkey V1 (denoted 'in vivo'), data generated from independent models matching the firing rates of culture 1 (denoted 'ind d'), and data generated from independent models with all equal firing rates (denoted 'ind h').

means that both the field parameter and the interaction parameters will be sensitive; however, while for the dissociated cultures the interaction parameters are sensitive due to the fact that through those interactions the key neurons exert their influence over the rest of the group, for the Poisson model these interactions are sensitive for the opposite reason – due to the fact that there are no significant interactions between units. In order to establish which of these scenarios held for the models fit to the *in vivo* data, as in Section 4.3.1, the average strength of interactions per neuron can be calculated and ranked by the neurons' sensitivity (Figure 5.11A).

Here, in contrast to the results from dissociated cultures (see Figure 4.16), it appears as though the principal unit in the group might exhibit lower, rather than higher average interaction. Confirming this intuition, panel B of the figure presents the ratio of leading unit interaction against the average of the rest, which in most cases falls below 1. This result indicates that in the *in vivo* recording the stiffest dimensions were determined by highly active but relatively isolated units, rather than highly connected ones. Thus the hypothesis of 'hub' neurons providing the stiff scaffolding for the sloppy groups to arrange around them does not find confirmation in the present results.

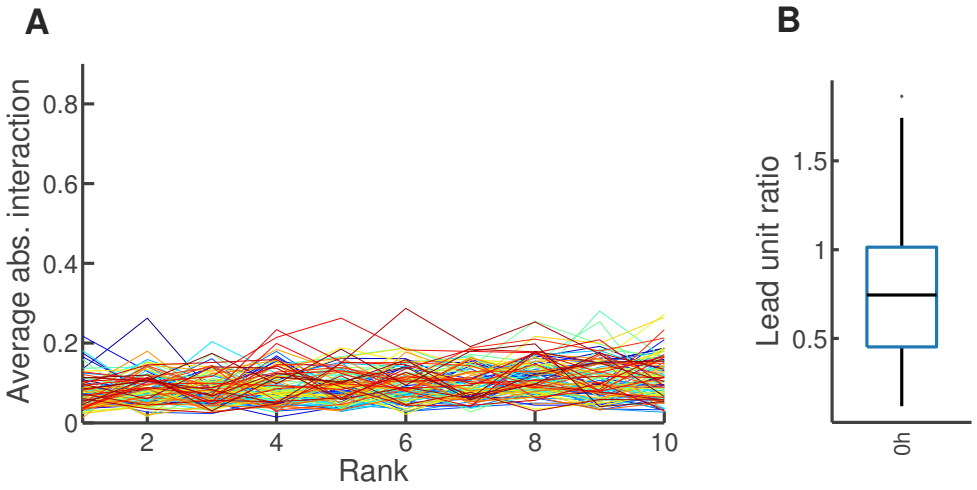

Figure 5.11: Average interaction strength per neuron in monkey V1. (A) Averaged for each unit, the strength of interaction parameter (absolute value); units are ranked by magnitude of average sensitivity; lines are color-coded according to the sparsity of the stiff dimension. (B) The distribution (across groups of neurons) of ratios of the average interaction strength of the most sensitive unit versus average interaction strengths of remaining neurons.

5.4 DISCUSSION

The current chapter of the thesis presents an attempt at extending some of the conclusions of the sloppiness investigation conducted *in vitro* into the domain of *in vivo* recordings. It was intended as a test of whether prerequisite conditions for the 'stability via sloppiness' hypothesis are met in the living brain. To that end, a recording of spontaneous activity from the primary visual cortex of the monkey (Chu et al., 2014a) was downloaded from a public neuroscience data repository and analysed exactly as the data from dissociated cultures. As transpired from this analysis, the key results could not be reliably replicated; first, a complicating factor of nonstationarity may have rendered the recording unsuitable for the framework of Ising model fitting; and second, although sloppiness and sparsity were observed, the structure of the stiff dimension was quite different from the one exhibited by the hippocampal dissociated neurons, casting doubt on the 'hub units' hypothesis. Although these results are not fully consistent with the *in vitro* findings, it is believed that they need not be treated as a negative proof, but rather as a challenge for more suitable and careful follow-up studies.

As introduced in Section 5.1, collective behaviour of *in vitro* and *in vivo* neurons can be vastly different (Corner, 2008); in particular, the ongoing spontaneous activity

in the brain exhibits a wider and more dynamic range of behaviours. The differences are already apparent from the most basic examination of activity, such as a comparison of the distributions of firing rates and correlations (see Figure 5.2). It is shown here that cells *in vivo* tend to fire more rapidly and exhibit a largely decreased level of synchrony as compared with the dissociated neurons grown on a microelectrode array, consistent with other studies (Hinard et al., 2012; Colombi et al., 2016). Also the results of modelling support these observations – although the fit quality of independent models was comparable across the different preparations used, the fits of the Ising models to the spontaneous activity in monkey V1 were better as compared to those of cultured neurons' (in fact near-perfect, as indicated by comparison with re-sampled re-fitted data), indicating a lower level of global synchrony (Figure 5.4). Such findings highlight the necessity for independent validation of hypotheses, when it comes to generalizing from *in vitro* results, since the conclusions might be particular to e.g. their synchronous bursting, which is not a representative state of *in vivo* networks.

However, the most notable discrepancy observed in behaviour was the different level of stationarity between the *in vitro* and *in vivo* recordings. Dissociated cultures, despite the nonstationary appearance of their raster plots, were in fact relatively stable from one half to the other, while neurons recorded from the cortex exhibited a notable change in their distributions of spiking patterns (see Figure 5.5). This highlights an ever-present issue with the analysis of *in vivo* data: the problem of attempting to understand a system that is in perpetual flux. As an example, consider the well-known problem of lack of reproducibility of spiking of a sensory neuron to a repeated stimulus (Brette, 2015): it has been shown that from trial to trial the responses of a given cell will differ. While this could be (and has been in the past) attributed to 'variability' or inherent noisiness of the neuron, a rather more straightforward explanation is that each trial, measured with respect to stimulus presentation, takes the spiking out of context: the context of the neurons' own inter-trial history, the context of the activity of its neighbouring units, the context of neuromodulation, etc. In fact, it has been shown that the variable state of a population can account for a large part of individual neurons' 'irreproducibility' (as discussed in (Brette, 2015)). In the present study, however, there is little clue as to the source of the 'irreproducibility' or even if that was a case of 'irreproducibility' or not. It is difficult to distinguish from activ-

ity alone whether the recorded units underwent a neuromodulatory change at some point during the recording (or perhaps only some of them, say only layer IV neurons); or whether, on the contrary, there was no change of general brain state but rather the observed nonstationarity was simply due to the fact that the network in idle mode of spontaneous activity cycles through available states and patterns in a complex time-dependent manner. Without that knowledge, however, no solid conclusions can be drawn from the analysis of model structure presented in this chapter.

Nevertheless, further analysis was performed under the assumption that the spiking patterns from both halves of the recording were, together, representative of the full repertoire. The Ising model fits were surprisingly accurate at capturing the distributions of neural code-words (see Figure 5.4) and, furthermore, revealed to possess a sloppy structure of parameter sensitivity akin to that observed in the *in vitro* data, as demonstrated in Figure 5.7. The identified stiff dimensions were also sparse; however, the source of that sparsity was fundamentally different from the one observed in dissociated cultures. While the dominant trend in the *in vitro* data was that the neural groups were lead by the more influential, better-connected units, here the trend was the opposite: the units identified as the most sensitive were characterized by lower average strength of connections than the rest. This suggests that the recorded population included some highly-active, but relatively isolated units. What could explain such a counterintuitive picture? One hypothesis is that this is an expression of a complex organizational structure of the primary visual cortex, where several distinct and sparsely interacting populations of neurons exist, forming a sort of mesh of parallel-processing threads. Under that scenario, the groups that were randomly sampled here do not form functional collectives; and to gain any further understanding of the above structures, a different group sampling strategy would need to be implemented in order to investigate them separately. It is interesting to speculate that perhaps such intertwined systems could operate under different principles each, both computationally and in terms of plasticity and stability (for example, should these separable populations carry different forms of information). However, this view conflicts with the results obtained by Yu et al. (2008), who also investigated the primary visual cortex employing Ising models. Not only did the authors find support for the existence of well-connected hub units but, furthermore, what transpired was that neurons with

sharp orientation tuning were in fact more likely to connect to neurons of different orientational preferences.

In summary, the hypothesised stability maintenance mechanism proposed in this work does not appear to be in operation in the particular dataset investigated in this chapter. However, it remains an open question whether this applies to *in vivo* conditions in general and no definite conclusions can be reached without further examination, taking care to exclude confounding factors and avoid the pitfalls listed above.

6

DISCUSSION

The work described in this thesis aimed to further the understanding of the interplay between the plasticity of individual cells and the collective stability observed in networks of neurons. This matter remains one of the outstanding questions in neuroscience (Marder and Goaillard, 2006; Turrigiano, 2011), as elaborated on in the first, introductory chapter of the present report. In the current work, it is proposed that a property inherent to many complex systems, *sloppiness* (Transtrum et al., 2015), might provide an explanation as to how the local homeostatic processes might be orchestrated in the network to achieve global stability. To test this 'stability via sloppiness' hypothesis, long-term observations of dissociated hippocampal cultures were coupled with model-based information-theoretic analysis of activity in groups of neurons. As reported in the previous chapters, the results of this investigation corroborate the posed hypothesis, suggesting that the global stability observed in neural ensembles can be attributed to the fact that the few hub-like leader units are homeostatically regulated, preserving the global functional structure and allowing local non-hub neurons to fluctuate without consequences. In addition, a tentative attempt at extending those results to *in vivo* conditions was performed, with inconclusive results. Herein the reliability and impact of those findings are discussed, their relation to other studies in the broader field of neuroscience, as well as suggested amendments, improvements and future prospects along this line of research.

6.1 LIMITATIONS

The methodological choices made come with certain necessary simplifying assumptions; here, the limitations imposed by those assumptions are examined.

First, the experiments conducted for this work employed dissociated hippocampal cultures as a model system of neural behaviour. This choice was motivated by the fact that *in vitro* preparations allow for reliable monitoring of hundreds of single

units over multiple days, exhibit the desired stability of behaviour and plasticity of single neurons (Minerbi et al., 2009; Slomowitz et al., 2015), and allow for a superior control of experimental conditions as compared with *in vivo* setting. While these are very favourable conditions for testing various scientific hypotheses, they come at the cost of a serious limitation – translatability to realistic conditions in the living brain. To reiterate the main points discussed in detail in Section 5.1: while cultures possess the tenets of brain circuitry, in that both exhibit small-world properties (Bettencourt et al., 2007; Yu et al., 2008; Schroeter et al., 2015), dissociated networks lack the complex organization of the brain, where intricate structure emerges from the interplay of environmental factors and biochemical cues from other brain areas (Corner, 2008). As a result, *in vivo* activity is much richer, varied and more dynamic than that of *in vitro* preparations; at best, it can be assumed that dissociated cultures reproduce something akin to the state of deep sleep (Hinard et al., 2012; Colombi et al., 2016). As an illustrative example, consider the work of Berkes et al. (2011), who demonstrate that the developmental changes that occur in the brain even after establishing complex structure result in significant re-configuration of functional representations. The authors of this study observe that as young ferrets are exposed to naturalistic stimuli, the pattern statistics of spontaneous activity in V1 progressively evolve to closely resemble that of the activity evoked by the stimuli (Berkes et al., 2011). What that entails for the present results is that even if brain circuits build upon small-worldness and sloppiness that appear innate to self-organizing circuits, the subsequent reshaping by development and experience can in principle completely overturn those properties and result in neural structures that behave according to entirely different principles. Thus, since the ultimate aim of any neuroscientific investigation is to gain a better understanding of the living brain, any conclusions drawn from studies performed in model systems and / or simplified conditions need to be re-tested and confirmed across a range of realistic conditions.

Secondly, the limitations of the pairwise maximum entropy model should be commented upon, as this is a fairly minimal, abstract model adopted from statistical physics, rather than a biologically-driven descriptive one, and that comes with a range of caveats. The Ising model has been shown to provide a surprisingly good description of the statistics for a variety of neural data (Schneidman et al., 2006; Shlens et al., 2006; Yu et al., 2008), matching well not only the first- and second-order

marginals, but also the entire spiking pattern distributions, suggesting that pairwise interactions are sufficient to explain collective correlated activity of groups of neurons. The analogy to a Hopfield network (Hopfield, 1982) makes it all the more appealing, as the parameters of the model appear intuitively understandable as properties akin to intrinsic excitability and synaptic strengths. However, as discussed in detail in Section 3.3.2, this mapping needs to be taken with caution. Firstly, as noted earlier, in the scatter-plot of field parameters presented in Figure 3.8 it could be observed that the model fits could at times yield inconsistent results across groups; a feature that was also observed in other studies (Schneidman et al., 2006). As discussed, this is due to the fact that – particularly in the present setting of sub-sampling the population – complications arising from unobserved neurons confound the interpretation of the field parameters. Specifically, these cannot be treated as reflecting the individual properties of the units they describe, since in order to well fit the marginals and spiking distributions the fields 'absorb' also the external influences that a given unit might be mediating.

Interestingly, when it comes to the matter of interactions, the results of studies such as that of Schneidman et al. (2006) and Yu et al. (2008) suggest that the inclusion of the complete circuit needs not be necessary for their unambiguous inference. In the first case, the authors observed that increasing the group size while keeping the core neurons fixed had little influence on the value of the inferred pairwise parameters (illustrated in Fig.4 b and c of Schneidman et al. (2006)). In turn, Yu et al. (2008) observed that estimated couplings for any given two neurons tended to be stable across models fitted to different groups containing the particular couple. This is presumably due to the small-world structure observed in their work (Yu et al., 2008), where only the hub neurons were widely connected and otherwise the connectivity was sparse. Since dissociated cultures are also found to exhibit such properties (Bettencourt et al., 2007; Schroeter et al., 2015) it is expected that similarly here the influence of unobserved units on the interaction parameters should not be prominent, particularly given the fact that the time bin was chosen so as to ensure the inclusion of only monosynaptic links in the analysis. However, there remains one important caveat not to be ignored: the fact that the definition of the maximum entropy distribution imposes a symmetric form of interactions, while synaptic contacts in real neuronal networks are directed and asymmetric. Therefore, also the interaction parameters of

the Ising model cannot be treated as reflecting the properties of individual pairwise connections.

This interpretational ambiguity brings with it two major limitations. First, while the analysis presented in this work technically allows to identify sensitive parameters and their changes, this translates in practical terms to identifying the sensitive units, but leaves doubts as to how this sensitivity is partitioned between synaptic connections from the neuron, synaptic connections to the neuron, and its intrinsic properties such as propensity to fire an action potential. And secondly, following from the above, this complicates the design of neuroscientific experiments aiming at independently testing the proposed hypothesis: it becomes unclear which of the above-mentioned variables should be monitored over time, and directly monitoring all of them simultaneously is not currently feasible (other than through employing alternative models and indirect inference, that is; but the ultimate test of stability in networks of living neurons should measure the biological properties that are under homeostatic control, not abstracted parameters).

The symmetry of interactional parameters mentioned above is also related to yet another limitation of the maximum entropy model – the fact that it is intended to describe stationary distributions only (i.e. since relationships in the temporal domain are not considered, there is no concept of causality, therefore no need for directionality). The fact that this strategy ignores the temporal statistics and thus prohibits any inference about the dynamics of neural activity, as shown by Tang et al. (2008), does not preclude useful insights into the functional structure of the network (Schneidman et al., 2006; Shlens et al., 2006; Yu et al., 2008) (as already discussed in detail in Section 5.3.2). However, this assumption of the maximum entropy model limits its applicability to cases where the recorded data can be verified to be stationary. This presents a notable challenge for attempts at applying the current methodology to *in vivo* conditions, as already apparent from the brief analysis reported in Chapter 5.

Another methodological choice that should be discussed here is the paradigm of sub-sampling the available population. As mentioned in Chapter 3, it is difficult to define what constitutes a functional circuit in the context of dissociated cultures; in fact, it is often unclear also *in vivo*. On the other hand, modelling an entire recorded population is both infeasible and likely pointless, as it is not realistic for the brain to employ such large ensembles simultaneously (consider that even with the ensem-

ble of 50 units the neural 'dictionary' of spiking patterns would consist of over a quadrillion entries; for comparison, the Oxford English dictionary contains approximately 600,000 words). Of course, many thousands of neurons become involved in any single task in the brain; these, however, are organized in hierarchical, modular structures. It was therefore assumed in this work that the 10-neuron groups are such putative modules, potentially overlapping, and that the extensive sampling would provide an idea of the average functional structure of such circuits. This number was a compromise between computational convenience and necessity, as explained in Appendix B; however, it cannot be denied that employing much larger groups might reduce interpretational ambiguity referenced a few paragraphs earlier.

Nevertheless, the construction of the Fisher Information Matrix of the Ising model, combined with its sloppy nature suggest that sloppiness should also hold for a range of larger groups. Recall that the FIM is in essence a matrix of covariances of observables, up to the fourth order, with respect to the model (see Section 4.2.2). Thus, matrices for larger groups will contain within them as subsets the FIMs of smaller groups (for illustration see Figure 6.1). Since the small matrices are all sparse (which implies anisotropy in parameter sensitivity), also the large matrices are expected to be sparse. Although it is not clear *a priori* whether the sloppiness would remain stable or progressively increase or decrease with size, it can be in fact tentatively tested. Since the modelled distribution is an approximation of the observed one, the matrix can equivalently be computed directly from the data; and this can be done for larger groups without the computational cost of fitting the model. In Figure 6.1 the results of such analysis are shown, comparing the Gini coefficients of eigenvalues of the FIM for groups of 20 neurons against groups of 10 neurons. As can be seen, not only do the larger groups also exhibit signatures of sloppiness, it appears in fact that it is more pronounced than in the smaller groups. It would thus seem that the results obtained in the present work are generalizable to larger ensembles.

However, the above considerations should be taken with caution. Employing data-calculated FIMs still relies on the assumption that the pairwise maximum entropy models provide a good fit to the observed distributions of spiking patterns. This fact was established here for the 10-neuron groups, but not for larger ensembles. And, as discussed in Chapter 3, model fit quality can not be reliably extrapolated to larger

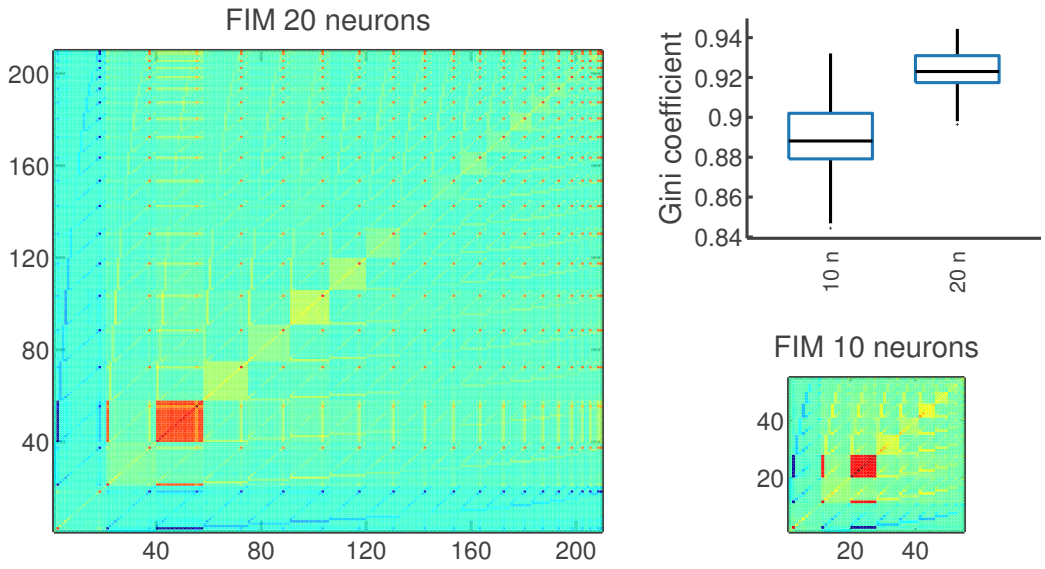

Figure 6.1: Illustration of the nested nature of Fisher Information Matrices of the Ising model and comparison of sloppiness of larger versus smaller groups (as measured with the Gini coefficient of the eigenvalue spectrum). The 10-neuron groups were subsets of the 20-neuron ones.

groups in the current setting (i.e. the perturbative regime, as discussed by Roudi et al. (2009b)).

This brings up the final point of the limitations of the maximum entropy approach: the question of whether the fit quality was sufficient to draw meaningful conclusions from model-based analysis. As stated in Section 3.2.2 the parameter adjustments continued until a relative convergence of marginals reached 10^{-5} (with the exception of some groups, where a lower convergence was considered sufficient if the discrepancy between predicted spikes did not exceed 2). That means essentially that the models were not a perfect fit to the data. Although the allowed discrepancy was small and could simply be attributed to noise, there is a possibility that the fitting error was not distributed evenly across the dimensions in model space. The main concern here would be if all of the fitting error clustered within the sloppy subspace, to the extent that the differences between time points could not be distinguished from differences in model fit quality at different time points.

Thus, additional analysis should always be performed in this framework, to confirm the significance of results. One angle that can be taken, and has been employed in this work, is to re-fit the models to data re-sampled from the fitted distributions,

to gain understanding of whether the sloppiness of the model fits might be conflated with sloppiness in the actual fitted networks. As can be seen from comparing Figure 4.21 with Figure 4.24 indeed the sloppy nature of error distribution decreases the size of the observed effect of changes over time; however, the parametric changes within the stiff subspace are still significantly smaller than those occurring along insensitive dimensions.

For a complementary approach, in Figure 6.2 the distribution of the fitting error across the eigenvectors is plotted, along with the distribution of eigenvalues (for reference as to the scales of the eigenvectors). As can be seen from this figure, and confirming the observations from the re-sampling re-fitting approach (see also Figure 4.23B for an illustrative picture of how fitting errors spread out with increasing rank of the eigenvectors), indeed the inaccuracies in fitting are not distributed evenly across the eigenspace. In plain words, the parameters corresponding to sloppy dimensions are more 'loosely' fitted than the stiff ones. However, even for the least significant ranks the fitting error remains smaller than the size of the sub-space, indicating that none of the dimensions is dominated by noise.

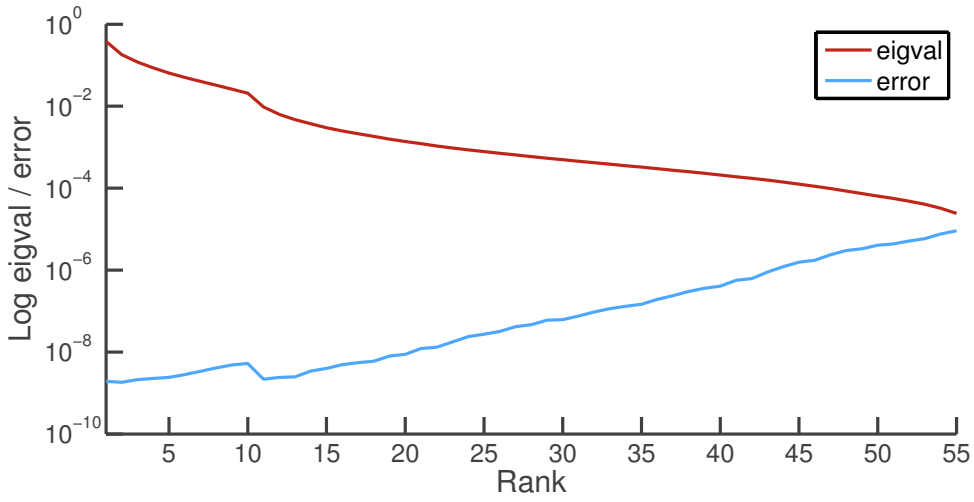

Figure 6.2: Illustration of the distribution of root square error of fitting among the eigenvectors of FIM decomposition. Plotted is the average root squared error across all groups of neurons from culture 1, baseline recording (blue); and average eigenvalue distribution across same groups (maroon) for reference.

6.2 RELATION TO OTHER WORK

The current study was, to the best of the knowledge of the author, the first to directly analyse sloppiness of models fitted to neural activity in groups of neurons. However, experimental research on homeostatic compensation and robustness has prompted theoretical considerations aligning with the current view (Marder and Goaillard, 2006), and other work in the field of computational neuroscience has addressed this problem from a complementary standpoint (Prinz et al., 2004). The experimental research in question demonstrates that in many neural systems there are a number of combinations of neural properties that can all result in the same behavioural and functional pattern (Marder and Goaillard, 2006). To take an already discussed example, in the stomatogastric ganglion of the crab it has been found that in different animals the same neurons contain different compositions of membrane ion channels. As a result, the individual ion conductances vary several-fold between animals; and yet, the neurons and the ganglions all exhibit the same behaviour (Golowasch et al., 1999; Marder and Goaillard, 2006). This has been further explored in a computational study by Prinz et al. (2004), who simulated over 20,000 instances of the stomatogastric ganglion model, to map the available parameter space with respect to resulting output of the model. As transpired from this research, 20% of examined models exhibited collective behaviour consistent with the experimentally observed pattern, and 2% could closely match the recorded data in 15 measures detailing the characteristic stomatogastric ganglion rhythm (Prinz et al., 2004). This finding suggests already that the parameter space of this model, and the corresponding neural circuit, might be sloppy. Interestingly, a closer inspection of the breadth of parameter values shown by the neurophysiologically realistic subclass of the models exhibited further indications of sloppiness. In particular, all but two of the conductances and synapse strengths could vary across the mapped range (Prinz et al., 2004). Thus, it would appear that the two synapse strengths were, in fact, the stiff parameters of the model.

As already acknowledged, the present results in themselves cannot be treated as evidence of sloppiness in the brain. However, the sloppiness hypothesis provides a compelling framework for the interpretation of a range of *in vivo* research, where stable functional representations emerge from populations of functionally and behaviourally variable individual neurons (Lever et al., 2002; Trachtenberg et al., 2002;

Carmena et al., 2005; Margolis et al., 2012; Ziv et al., 2013). Another interesting study relevant in this context is the very first study mentioned in the present report, wherein monkeys were trained to operate a brain-computer interface (Ganguly and Carmena, 2009). In this work, the animals were first trained on a fixed decoder to perform a reaching task, until they attained expert performance. The neural population controlling the decoder performed in a stable manner; furthermore, the authors find that also individual neurons contributing to the ensemble were stable both in activity and function (Ganguly and Carmena, 2009). This is in contrast to work discussed in the introduction, where stable representations were found to consist of neurons with fluctuating membership or tuning properties. Note, however, that the recorded sample of neurons in the above study was relatively small and confined to a small area of the cortex; thus it likely encompasses only a subset of all task-relevant neurons (potentially read-out neurons) while a much larger number of cells is involved downstream, processing information and converging on the chosen units.

As the monkeys were trained again on a different instance of the task, but employing the same neurons for decoding, an interesting phenomenon occurred. The animals were able to acquire expert performance on the second task, without interfering with the first one; and crucially, the functional properties of only some of the neurons were re-mapped between tasks, while a portion of the population remained stable between the two contexts (Ganguly and Carmena, 2009). A potential explanation for such results is that the functional representations of motor tasks rely on overlapping neural ensembles that share stable units encoding widely generalizable properties. These units and their synapses would be expected to be stiff and remain largely stable (that is, respond the same across tasks), while the neurons encoding less generalizable traits would be more amenable to plasticity. Thus, the downstream learning-related changes would induce re-mapping in this sloppy subspace, resulting in a differentiation of behaviour of the less generalizable units between contexts. Such structure of complex representations in the cortex has been in fact postulated by Hung et al. (2014), with invariant units encoding generalizable information and more variable and heterogeneous population providing fine-grained details. In this study, also performed on monkeys (under light anaesthesia), the authors find that the output layers of the IT cortex tend to contain correlated 'chorister' neurons responding to broad categories, while the input layers of IT consist mostly of less correlated

'soloist' cells carrying less generalizable information (Hung et al., 2014). Considering the above studies, it is tempting to hypothesise that stable functional representation should rely on stiff, stable neurons, encoding generalizable properties of the world, while sloppy cells responsive to more detailed information could undergo learning to encode novel stimuli belonging to a known broader class but distinguishable by less common characteristics. This is broadly in line with the hypothesis, postulated by Buzsáki and Mizuseki (2014), of the preconfigured log-dynamic brain, where the minority of fast-firing strongly connected neurons are linked to high-probability network states, which over development become associated with common or generalizable stimuli in the world; at the same time the less active and less connected majority of cells retain high plasticity and are able to fine tune the low-probability network states to particular stimuli. However, considerable further work is needed to address those questions adequately and to understand the relationship between stability and structure of functional representations in the cortex.

In their paper proposing the hypothesis of the preconfigured brain, Buzsáki and Mizuseki (2014) distinguish between the log-normal distributions observed in various types of data (such as firing rates, synapse strengths or even sizes of place fields in the hippocampus) and the power law distributions characteristic to the scale-free critical networks (Beggs and Timme, 2012). This is pertinent here, since clues of both log-normality (in the firing rate distributions) and scale-free behaviour (suggested by sloppiness, as discussed below) can be observed in the present results. In general, inferring criticality is wrought with difficulties and the observation of power-law distributions does not guarantee scale-free behaviour of the system; nor is lack of such observations in neural data precluding criticality (for an insightful discussion see Beggs and Timme (2012)). In view of the complexity of the matter, it is difficult to draw firm conclusions from the present results. However, it is interesting to note that sloppiness is linked to critical behaviour, as exhibited in the example Ising model discussed by Machta et al. (2013). The very definition of critical behaviour is that the system progressively becomes governed by a single parameter, the critical parameter, regardless of the microscopic ones; which is conceptually equivalent to increasing sloppiness. And, as noted earlier, in contrast to the parameter space anisotropy observed by Machta et al. (2013), where it was induced by scale-coarsening, here it was observed to emerge naturally. Furthermore, the fact that sloppiness appears to

increase with group size (as shown in Figure 6.1) is resemblant of reports of other researchers, wherein Ising models fit to data from retinal recordings were also shown to move closer to the critical point with increasing system size (Tkacik et al., 2006). Altogether it appears that the fitted pairwise maximum entropy models were near-critical. However, it is the view here that the notion of criticality in the brain should be treated with caution; while being poised close to a critical point might carry certain computational advantages (Shew et al., 2011), the evidence in the field is far from conclusive (Beggs and Timme, 2012).

6.3 FUTURE DIRECTIONS

As discussed in Section 6.1 the validity of presented results hinges upon a number of limiting assumptions, having to do with experimental conditions and methodological choices made. A natural extension to the current work, therefore, would be to test the proposed hypothesis under alternative conditions; and through complementary analysis framework.

Since the findings reported here are applicable to a regime of neural activity resemblant of the behaviour of cortical cells in deep sleep (Hinard et al., 2012; Colombi et al., 2016) (see Section 5.1 for details), wherein characteristic synchronous bursts are prevalent, the first proposed extension of present results would be to employ data of a more asynchronous, wake-like nature. And as *in vivo* recordings currently present a considerable methodological challenge, a viable alternative would be to employ suitably modified *in vitro* data. As observed by Hinard et al. (2012), dissociated cultures exhibit not only electrophysiological, but also molecular and metabolic signatures of a deep sleep-like state; which prompts the idea that application of appropriate neuromodulators might switch the *modus operandi* to a 'waking mode' of activity. Indeed, Hinard et al. (2012) and Colombi et al. (2016) demonstrate that this is possible, resulting in modulation of activity at all levels, including gene expression and spiking behaviour of studied neurons. With the analysis framework already in place, this intervention presents an excellent testing ground for the hypothesis proposed here: it combines a desirable *in vivo*-like global behaviour regime with the ease of access, manipulation and long-term monitoring afforded by the *in vitro* approach.

Should such tests prove successful, the natural next step would be to establish an experimental setting allowing to employ sloppiness analysis in the living brain. As shown in Chapter 5 this is, however, not a trivial undertaking, considering the complications of uneven coverage and biased sampling (Shoham et al., 2006) *in vivo*, and given the nonstationary nature of neural activity. Addressing the former of these issues is a challenging task, testing the limits of available technology. The best recording method from the point of view of reliability of day-to-day monitoring and faithful representation of the population activity distribution is two-photon calcium imaging (Lütcke et al., 2013). However, the temporal resolution of this signal is poor (on the scale of 100ms, which is 20 times the resolution of binning employed in the present work), precluding approaches similar to the present one. This method would be thus more suited towards experimental validation of testable predictions, rather than direct sloppiness analysis. The alternative to calcium imaging is direct recordings of extracellular potential through implantable multielectrode arrays. This allows for excellent time resolution and identification of single spikes; however, requirements of spike sorting algorithms tend to make it biased towards highly active units (Lütcke et al., 2013). Therefore, effort should be made to obtain long recordings from large numbers of cells, to ensure representation of a wide range of firing rates under examination, since it is likely that the sloppy neurons will also tend to exhibit lower activity. While challenging, this could be achievable with systems such as that recently introduced by Schwarz et al. (2014), where chronic imaging from close to 500 neurons was reported.

In order to solve the latter of the aforementioned issues, i.e. the problem of nonstationarity of brain activity, it would be beneficial to use much longer recordings than the time scales used here, and accompany these with additional monitoring or analysis methods. As explained in Section 5.3.2, the core of the problem is that if a single recording session comprises two different modes of activity, the statistics derived from such data would conflate the two states, ending up representing neither faithfully. For an illustrative example as to how this could be overcome, consider that different stages of sleep, characterized by different dynamics of neural activity and neuromodulatory environment (Hobson et al., 2000), are associated also with distinct electroencephalographic profiles (Rechtschaffen and Kales, 1968). This association between micro- and macro-scale activity profiles could be exploited to segregate periods

of multielectrode recordings as belonging to different mental states by identification through standard electroencephalography procedures.

In the context of discussing the limiting assumption of stationarity inherent in the maximum entropy model, it bears reminding that an important aspect of neural activity is the fact that spiking patterns follow one another in a meaningful temporal order. In fact, various forms of temporal dependencies have been implicated as crucial to neural coding (Shahaf et al., 2008; Brette, 2015). Thus, although understanding the stability of the code-words of neural dictionary is an important step towards solving the plasticity vs. stability dilemma, these results are not complete without addressing also the stability of temporal sequences of spike patterns. In order to tackle this issue, an alternative approach to the pairwise maximum entropy model is required. One viable option would be to employ a form of a statistical model capable of capturing the properties of spatio-temporal patterns; a natural choice here would be an asynchronous kinetic Ising model (Zeng et al., 2013), which is a generalization of the maximum entropy model to non-equilibrium conditions. This model also describes the probabilities with a distribution from an exponential family, and the Fisher Information Matrix can be directly calculated (Zeng et al., 2013), lending it amenable to sloppiness analysis.

Another direction for future research that is crucial for validation of biological relevance of the present results would be to combine modelling and sloppiness analysis with a direct method of monitoring functional connectivity. As discussed earlier, it is unclear how the functional parameters of the Ising models map onto biological connectivity. It is therefore important to examine whether synapses corresponding to presumably sloppy parameters exhibit larger fluctuations than those of the putative stiff parameters. This could be implemented in dissociated cultures in analogous manner to Slomowitz et al. (2015), where monitoring many units through multielectrode arrays was combined with single-cell recordings, thus allowing for analyses of evoked post-synaptic potentials. Only through such research it can be reliably established whether homeostatic plasticity tends to operate on the few sensitive parameters.

A further interesting extension of the present results would be to investigate in more detail the relationship between the structure of the stiff dimensions and the properties of the corresponding neurons. A particularly appealing hypothesis is that the sensitive leader neurons identified here correspond to hubs in the small-world

networks observed in self-organized ensembles of neurons (Bettencourt et al., 2007; Yu et al., 2008; Schroeter et al., 2015). As has been observed by Schroeter et al. (2015), hub units in dissociated hippocampal cultures form early on in the development of the network and tend to persist throughout weeks of observation; this stability suggests a possible link to the hypothesis advanced in the present work. Therefore, although it has been identified here that the sensitive neurons are more strongly connected and more active than the insensitive ones, there are a number of additional measures that could verify this association. For example, a similar analysis to the one performed by Schroeter et al. (2015) could be performed, yielding parameters such as nodal efficiency, clustering coefficient, betweenness centrality, hub score etc.

Finally, it is hoped that the framework of sloppy research will be more widely adopted within the computational neuroscience community, in applications to other models of neural activity. Although the pairwise maximum entropy approach has its merits, simple alternatives such as the Dichotomised Gaussian (Yu et al., 2011) appear more widely applicable and more easily scalable. Furthermore, since the sloppy theory framework was designed with the aim to aid in understanding complex, multiparameter models based on differential equations (Brown and Sethna, 2003; Transtrum et al., 2015), it could also be applied to a range of more biophysically accurate neuroscientific models such as Hodgkin-Huxley description of simple circuits (Golowasch et al., 1999). It could also be employed in purely theoretical analyses of properties of artificial neuronal networks such as e.g. the Izhikevich model (Izhikevich, 2003). It would be highly interesting to see whether stability analysis conducted via such alternative methods would corroborate the present results; and if so, whether these results generalize also to *in vivo* activity.

6.4 CONCLUSIONS

The brain is a beautifully intricate biological machine, not only orchestrating the behaviour of organisms, but endowing them with the ability to learn new functions and adapt to a changing environment. This seems an extraordinary feat, when considering the web of intertwined processes that need to remain reliably stable while plasticity takes place. And indeed it might seem so from the point of view of a researcher trying to unravel the design of such a complex and adaptable system. This

design, after all, emerged from millions of years of evolution. It is perhaps then not surprising to find that the emergent nature of complex activity manifests also in the present results. The sloppiness observed in the activity of hippocampal cultures demonstrates how stability can naturally arise in systems with few sensitive parameters and many insensitive ones, mirroring findings from other areas of biology. It might then be, as already suggested (Transtrum et al., 2015), that sloppiness is inherent to life as we know it, since it is the simplest emergent solution to evolutionary challenges.

A

SUPPLEMENTARY INFORMATION: ENTROPY ESTIMATION

A.0.1 Introduction to the entropy estimation problem

In the main body of the thesis it is stated that entropy estimation from limited data is subject not only to random variability errors but also to systematic biases. The issue of random variability error is one easily understood and demonstrated: when generating limited-length data from any probability distribution, it is intuitively clear that the random process of drawing samples will turn out, at least at times, counts that by chance over- or underestimate the true probabilities. For example let us take the re-sampled data used in this work, where the 10x180000 matrices of spikes were generated by assigning one out of 1024 possible 10-neuron state vectors to a time bin if a randomly generated number fell within a prescribed range of values (the probability of which was dictated by the distribution re-sampled from). As shown in Figure A.1, different instantiations of re-sampling result in different, small deviations from the true probability distribution. The resulting value of entropy will therefore be different from the true one, regardless of the estimation method; furthermore, in each realization of sampling this difference can change both in magnitude and sign, as in the example presented in Figure A.1.

It is less intuitive, perhaps, that repeated re-sampling from that same distribution, when averaged over many runs of generating artificial data, turns out to lead to a systematic difference from the true entropy (see Figure A.2). This is not any more due to random chance, but rather to the fact that the number of time bins in the data limits the resolution with which probability can be represented. The resulting systematic difference in entropy is called the limited sampling bias, and it crucially depends on both the estimator used and the data in question (note that the limited sampling bias affects also probability-dependant measures other than entropy; however, for simplicity the focus here is on entropy only).

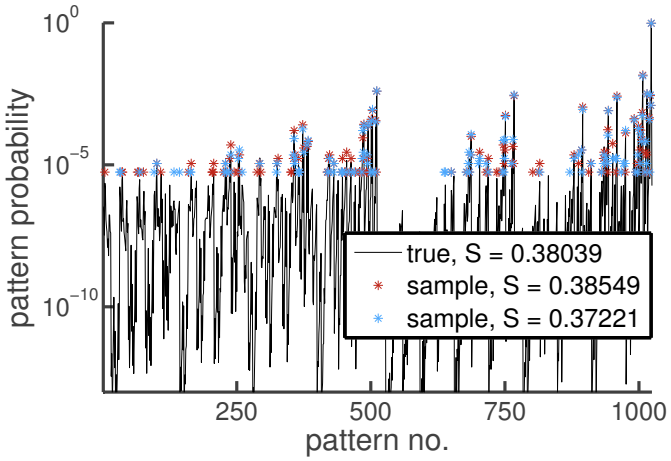

Figure A.1: Illustration of the effect of random factors in probability and entropy estimation from limited data. Black line denotes the true probability distribution, derived from a pairwise maximum entropy model. Blue and maroon stars mark two re-sampled probability distributions, and entropy is then calculated via the naive estimator.

To better understand the complexity of the issue, let us turn to a very pertinent example from a recent paper by Macke et al. (2013). There the authors analyse the properties of an entropy estimator derived from the pairwise maximum entropy model. In this approach, in stead of using the naive (or plug-in) estimator

$$S_{data} = -\sum_i \frac{c_i}{K} \log \frac{c_i}{K}, \tag{A.1}$$

where c_i denotes the counts of state i and K denotes the number of bins in the data, one fits a pairwise maximum entropy model and then treats the entropy of obtained model as the estimate. The authors derive a general expression for the approximate bias of this estimator under any type of data (see Equation 11 in Macke et al. (2013)). They subsequently show that for data that is generated by a pairwise maximum entropy model, this general expression simplifies to a closed form solution

$$b = -\frac{m}{2 \cdot K}, \tag{A.2}$$

where m is the number of constraints in the model. They thus pose the problem in a framework that is useful to adopt here: there is an irreducible amount of bias even in the scenario where the model assumed for the entropy estimator is consistent with

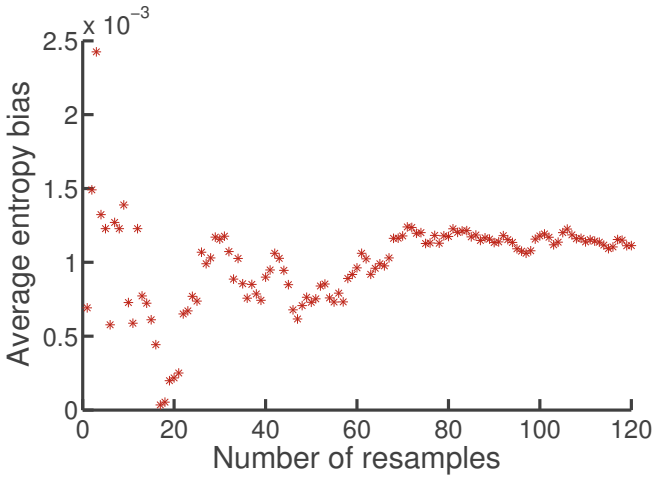

Figure A.2: Illustration of the effect of systematic factors in entropy estimation from limited data. Average bias (difference between the true and the estimated entropy) is plotted as a function of the number of re-sampling instances (i.e. generated datasets); as the number of re-samples increases the average bias converges to a steady value. (Note this is for illustration purposes only; the presented number of re-samples is not necessarily sufficient to converge to the analytically calculated bias)

the one generating the data; and there is an additional amount of bias whenever there is a model misspecification.

For an illustration let us look again at the re-sampled data. For each artificial matrix the entropy can be calculated by the naive estimator, and compared to the true one, as plotted in Figure A.3. Alternatively, a pairwise maximum entropy model can be re-fit to that data, and the entropy of the model distribution used as an estimator. Both methods on average show a bias, however, the maximum entropy estimator performs better than the naive one and that is because it's assumptions are consistent with the model that generated the data.

A.0.2 Motivation behind the re-sampling re-fitting approach

Therefore, whenever using entropy estimators, three factors should be considered: a) random variability, b) limited sampling bias, and c) model misspecification. What complicates the matter further here is the fact that in the multi-information ratio used in this work there are several entropies entering the calculation. First, there are the entropies of the independent and Ising models fit to the data. As already explained above, even if the data was generated by a pairwise maximum entropy

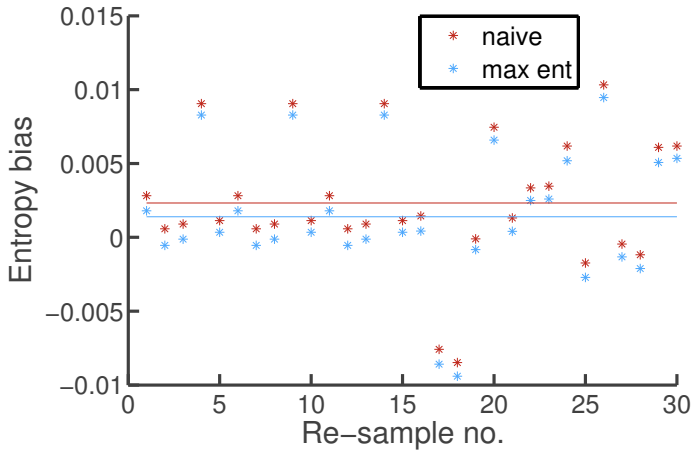

Figure A.3: Illustration of the difference in the systematic bias between two estimators, naive (plug-in) and maximum entropy model-derived, for data generated by maximum entropy model; the estimator consistent with the data exhibits a lower bias than the naive estimator.

model, using these entropies would lead to a multi-information ratio lower than 1, due to factor b). Thus on one hand the model-related entropy biases result in possible underestimation of the true MIR, although these effects are expected to be small based on Equation A.2. Second, there is the data entropy, which is calculated by the naive estimator from the observed probability distribution and thus subject to both b) and c). This, on the other hand, results in systematic biases that overestimate the MIR. Finally, the random variability factor a) is capable of driving the multi-information ratio in either direction, depending on chance only. Thus, the problem of calculating multi-information ratio is faced with a complicated interplay of factors.

What is proposed in this work to simplify this rather confounding picture is a re-sampling re-fitting approach. In essence, by calculating the multi-information ratio for models re-fitted to re-sampled data, an estimate is constructed of the combined effect of all biases and variability on the model evaluation measure: the re-sampling imitates the random error in generating data from a model; re-fitting imitates the fitting and provides a computational estimate of the limited sampling bias affecting maximum entropy models; and finally the naive entropy estimator applied to artificial re-sampled data imitates the combined effects of b) and c) on the experimental data. The MIR for models fit to experimental data can then be compared to the re-sampled MIR, and any difference between them can be attributed to failures of the model in accurate description of the data, rather than the biases.

It should be noted that this approach tacitly assumes that the observed data is similar enough to artificial data so that the b) and c) affecting the naive estimator are similar for both. This simplifying assertion is, however, not in its essence much different from e.g. using the Miller-Maddow bias correction

$$b = -\frac{2^N - 1}{2 \cdot K}, \tag{A.3}$$

which was derived for a multinomial distribution; these approaches are not different in the sense that both necessarily rely on *a priori* assumptions regarding underlying distributions.

A.0.3 Results and discussion of entropy estimation

The re-sampling re-fitting approach, as described in the main text, resulted in generating 30 artificial spike train matrices per set of neurons. Pairwise maximum entropy models were then fit to each artificial dataset. Thus, for each out of 495 sets of neurons there were 30 plug-in estimates of entropy and 30 maximum entropy-derived estimates.

In Figure A.4 the results of these simulations are presented, with the total relative bias of each re-sample plotted as a function of true entropy. In addition also plotted are the averages across re-samples for each set of neurons (black) and the analytically predicted limited sampling bias (yellow). First, for both estimators the combined effects of various biases appear to change the entropy value by at most 6%. Furthermore, as evident from a comparison with the average values per set of neurons (indicated in black in Figure A.4), the dominant contribution to errors in entropy calculations comes from random, rather than systematic factors. Thus, it appears that the effects of entropy estimation on the calculation of MIR are likely to be minor, and dominated by random variability.

Second, there is a trend of diminishing impact of errors with increasing entropy, as evidenced by the funnel shape of scatter-plots. Interestingly, this is most evident when examining the spread of variability (color) rather than the systematic effects (black). Such a trend would be expected of systematic factors, since they have fixed value, and therefore will affect most (in relative terms) the sets of neurons with low

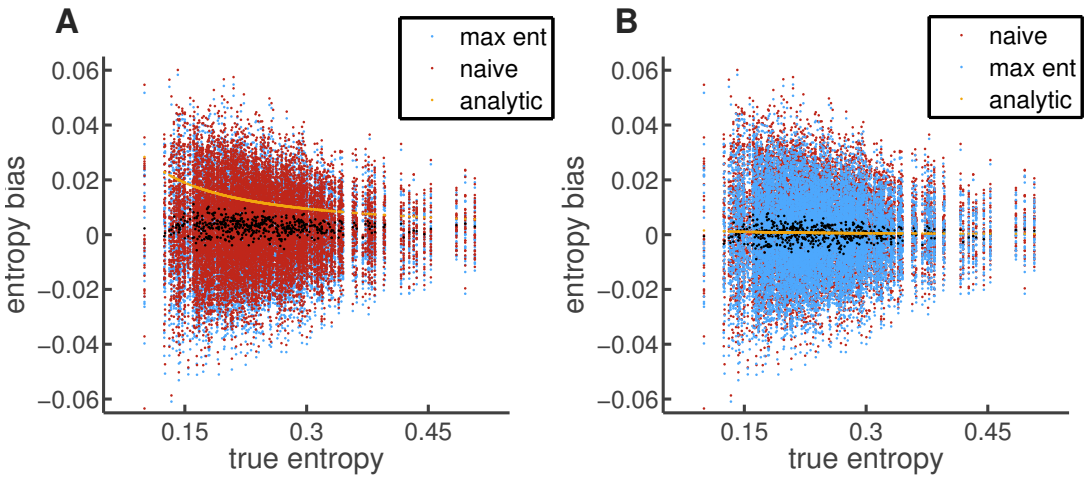

Figure A.4: Illustration of biases in entropy estimation obtained from re-sample re-fit simulations of maximum entropy models. (A) Relative entropy bias as a function of entropy for the two estimators used, with the plug-in results overlaid on the maximum entropy model-derived; also plotted is the Miller-Maddow analytical prediction of the bias (yellow). (B) Same as in A, but with model-derived bias overlaid on the plug-in results; also plotted is the analytical prediction for the maximum entropy estimator (yellow).

entropy (see for illustration the yellow marks in Figure A.4A denoting the predicted systematic bias). The fact that this trend is not evident in the average values per set is due to the fact that for the modest 30 re-samples used here the averaged value of bias does not converge to its predicted asymptotic value.

The above observation of the lack of convergence is of particular import here. In the case of the maximum entropy estimator (Figure A.4B), where the asymptotic bias is very small, it highlights the dominance of random variability over the systematic issues. In the case of the naive estimator, on the other hand, the asymptotic bias is higher than the averaged values. This suggests that the Miller-Maddow correction used here overestimates the impact of systematic biases in this type of data (since it was derived for a multinomial distribution, and not a maximum entropy one). Thus, while the naive estimator is flawed and subject to systematic errors, it appears that for the type of data used in the simulations it does not fail as badly as predicted by the correction of Equation A.3.

Altogether, the re-sampling re-fitting approach allowed to uncover the dominance of random variability and the pitfalls of using correction terms such as the Miller-Maddow.

B

SUPPLEMENTARY INFORMATION: INCREASING THE GROUP SIZE

B.0.4 Introduction to the group size problem

The analysis reported in this short supplement was inspired by the notion introduced in Section 3.3.2 that the interpretational ambiguity of model parameters (when applied to randomly sampled groups) could in principle be overcome by including the complete circuit in the inference. Since the data here comes from densely sampled dissociated cultures, it might be assumed that there is sufficient information in these recordings to infer the properties and connectivity of all active neurons, if only all of them are included in the model. However, as discussed already in Section 3.2.3, it is computationally infeasible to fit an Ising model to an entire population of units from an *in vitro* recording; and thus the present type of analysis is necessarily constrained to employ sub-sampled groups. Yet, the question remains whether increasing the group size might still be beneficial for analysis: if not amending, then at least ameliorating the ambiguity of parameters.

From a practical standpoint, there are two considerations to take into account in this matter. Firstly, with increasing number of neurons the reliability of the assessment of model fit quality becomes itself less reliable, complicating the interpretation of the findings. Recall that the number of observable patterns scales as 2^N; therefore, adding a single unit to a group doubles the size of the support of the pattern probability distribution. Thus, to obtain a matching reliability of the observed distribution estimate, the recording time should be (at the very least) doubled as well; and without increased observation time, it becomes difficult to dissociate the effects of poor model fit from the effects of limited sample size. Furthermore, and importantly, the computational cost of the maximum entropy model fitting procedure grows exponentially with each extra unit added to a set, which is a steep price to pay for an increase in information that is difficult to estimate, but certainly sub-exponential.

Here, as a tentative exploration of the costs and benefits of increasing the group size, a brief analysis of the topic is reported.

B.0.5 Methods

The baseline recording from culture 1 was used for the present analysis. As in the main body of the thesis, groups were randomly chosen from the available population; here it was 60 groups of 16 units, resulting in each neuron being chosen approx. twice (as opposed to 10 in the main work; this was due to computational constraints). Groups of smaller size were constructed by choosing only a portion of the original 16-unit ensembles; group sizes used were 4,6,8,10,12,14 (with the core 4 units preserved across group sizes). Then, maximum entropy models were fitted to each of the groups at each size step, with the same algorithm as in the main work, and same fitting criteria (see Section 3.2.2). Measures of pattern distribution fit quality were also the same as used in the main work, i.e. the multi-information ratio and Shannon-Jensen divergence (see Section 3.2.4). Finally, in order to obtain an idea of the increased impact of the limited sampling bias with increased group size, for each of the fitted models a single run of re-sampling was performed, and MIR and JSD were calculated between the model and re-sampled spiking pattern distributions.

B.0.6 Results and tentative implications

First, in Figure B.1 the estimates of model fit quality for each of the chosen groups and each chosen size are reported. Note that with increasing number of neurons the average multi-information grows but MIR decreases (panel A of the figure). The growth in multi-information is expected, since with increasing group size the failure of the independent model becomes more prominent (that is, more units are included and the synchrony between them can become apparent). Also some decrease of multi-information ratio can be expected, due to the limited sampling bias; however, it cannot be excluded that as the group size increases, the failure of the Ising model also becomes more obvious as more higher-order correlations are likely to be included.

For further examination of this possibility, panel B of Figure B.1 depicts the increase in the Shannon-Jensen divergence between the modelled and observed spiking pat-

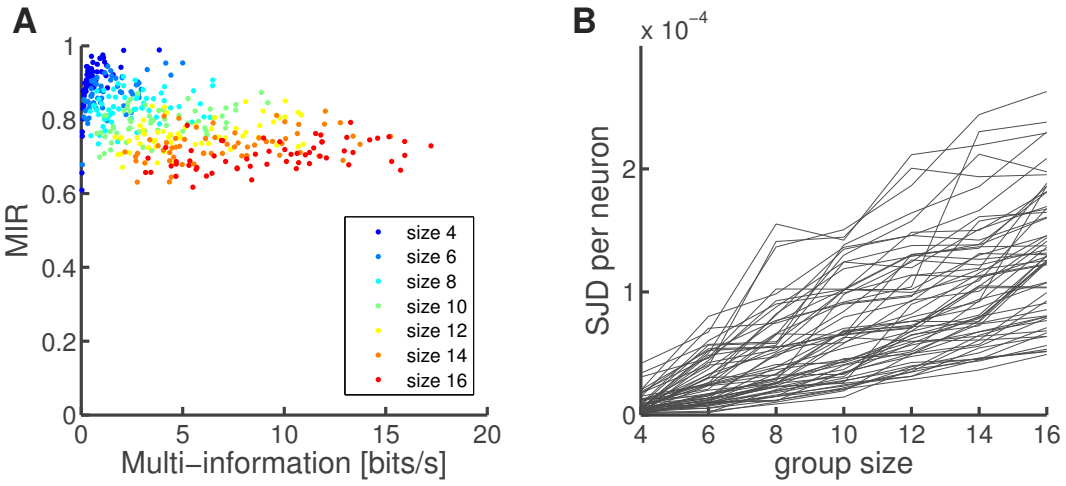

Figure B.1: Illustration of the change in model fit evaluation measures with increasing size of the fitted group. (A) Multi-information ratio for 60 randomly chosen groups of units of increasing size (color-code by group size). (B) Jensen-Shannon divergence per neuron (between the modelled and observed pattern distributions of firing patterns) as a function of group size, for each of the 60 randomly chosen groups of units.

tern distributions. Note that what is plotted is not JSD itself, but JSD per unit; this means that the divergence of model fit increases super-linearly with the number of neurons added. Again, such results could stem either from the limited sampling bias or higher-order effects becoming prominent with inclusion of extra cells (or, in fact, a combination of both).

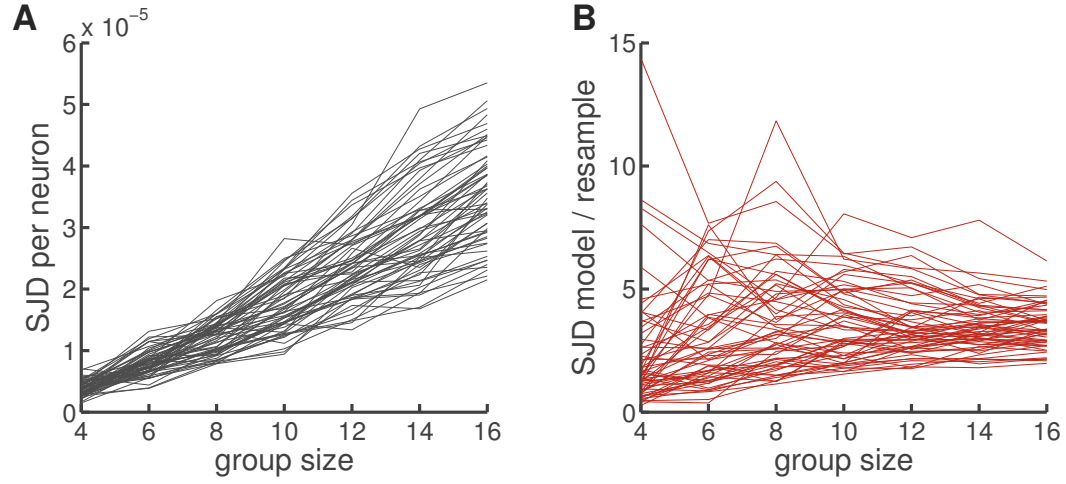

Figure B.2: Illustration of the change in model fit evaluation measures with increasing sampling bias. (A) Jensen-Shannon divergence per neuron (between the modelled and re-sampled from model distributions of firing patterns) as a function of group size, for the 60 randomly chosen groups of units. (B) The ratio between the model-data Jensen-Shannon divergence and model-resample Jensen-Shannon divergence, as a function of group size.

To get an intuition of how limited sampling bias behaves with increasing group size, in Figure B.2A a similar plot was constructed for the data re-sampled from the fitted models. Perhaps not surprisingly, also here the divergence grows super-linearly (recall the exponential growth of the number of possible patterns); and in fact with what appears to be a steeper slope than that in Figure B.1B – this, however, is not very indicative of model fit quality, as the re-sampling was performed only once, and few groups in total were employed. Furthermore, notice the difference in scales on the y-axis between the two figures; the match between distributions is worse for real data than the re-sampled one.

For a more direct comparison, in panel B of Figure B.2 a ratio of Jensen-Shannon divergences of model-data to model-re-sample is plotted (per each group separately). This uncovers an interesting effect: with a growing number of neurons, the ratio appears to saturate at an average level of approx. 4, and the variability between groups decreases. This indicates that although the Ising model can not faithfully capture the complete compound correlation structure of the data (noted and discussed already in Section 3.3.1), it would appear that the problem does not worsen with group size. Although it cannot be extrapolated to even larger groups without testing, it is an encouraging result, suggesting that model fit quality might not present an issue for analysis of larger groups (at least for the type of data employed here).

This leads us to the second practical consideration regarding group size, i.e. increase in informativeness versus computational costs. In Figure B.3 scatterplots of model parameters versus neuron activity properties are presented, in analogous manner to Figure 3.8. Notice a largely similar picture presented by both versions of the fitting, particularly the clearly nonlinear mapping between Pearson correlation coefficients and interactions, with a characteristic trend of saturation towards higher firing rates correlations. Importantly, note the fact that also here, as in the case of groups of 10, there are signs of field parameter ambiguity, that is, cases where a single neuron has two different field parameters, depending on which group it belongs to (a particularly striking example is the unit with firing rate just above 3Hz, circled in blue, which in one case attains the field value of below 2, and in another nearly 4.5).

Of course, increasing the number of units from 10 to 16 is hardly expected to fix the issue of potentially missing out an influential neuron from a group, since the whole population consists of hundreds of cells; the point here though is that this appears

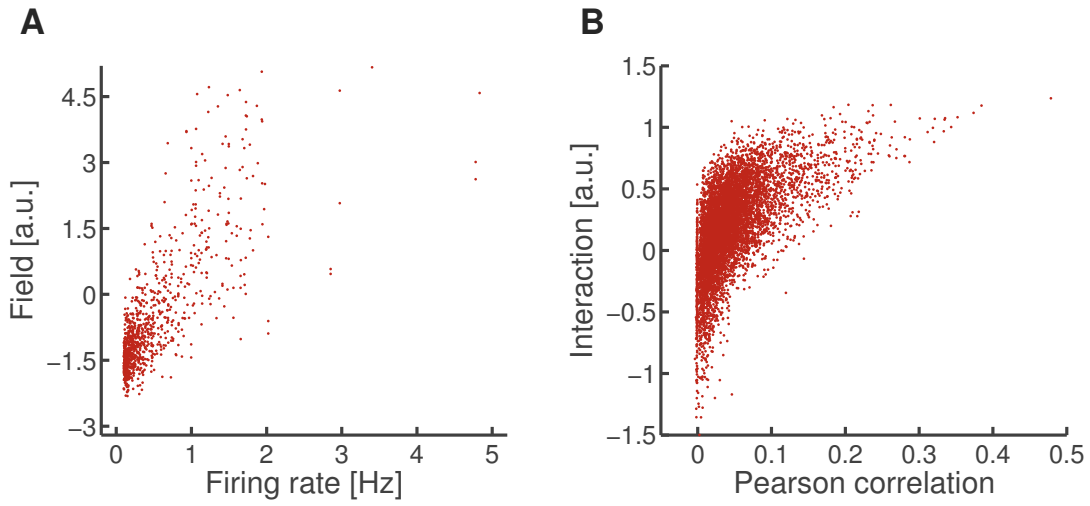

Figure B.3: The distributions of the parameters of fitted models for the largest used group size, 16. (A) Scatterplot of model field parameters against the corresponding firing rates of neurons, pooled across all sets from culture 1, baseline recording. (B) Scatterplot of model interaction parameters against the corresponding Pearson correlation coefficients of pairs of neurons, pooled across all sets from culture 1, baseline recording.

to offer little improvement in the interpretability of parameters – they still need to be understood in the functional context of the group they belong to – and still much larger numbers would need to be considered for novel insights. The computational cost for this increase is, however, prohibitive, since with each added unit, the time it takes to fit a single group of neurons doubles. To give the reader an idea of how this would practically affect the work performed here, consider that fitting the approx. 500 groups to the baseline recording of culture 1 took a little over 48 hours. If each neuron were to be used approx. 10 times in 16-unit groups, that would require about 320 sets, each one taking 2^6 times the computing time of the 10 neuron set, resulting in about 81 days of computation for a single recording.

BIBLIOGRAPHY

Ackley, D. H., Hinton, G. E., and Sejnowski, T. J. (1985). A learning algorithm for boltzmann machines*. *Cognitive Science*, 9(1):147–169. (Cited on page 55.)

Amari, S.-i. and Nagaoka, H. (2007). *Methods of information geometry*, volume 191. American Mathematical Soc. (Cited on pages 92, 93, 94, and 135.)

Amin, H., Maccione, A., Zordan, S., Nieus, T., and Berdondini, L. (2015). High-density meas reveal lognormal firing patterns in neuronal networks for short and long term recordings. In *Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on*, pages 1000–1003. IEEE. (Cited on page 44.)

Arnold, F. J., Hofmann, F., Bengtson, C. P., Wittmann, M., Vanhoutte, P., and Bading, H. (2005). Microelectrode array recordings of cultured hippocampal networks reveal a simple model for transcription and protein synthesis-dependent plasticity. *The Journal of physiology*, 564(1):3–19. (Cited on page 9.)

Barral, J. and Reyes, A. D. (2016). Synaptic scaling rule preserves excitatory-inhibitory balance and salient neuronal network dynamics. *Nature Neuroscience*, 19(12):1690–1696. (Cited on page 2.)

Beggs, J. M. and Timme, N. (2012). Being critical of criticality in the brain. *Frontiers in physiology*, 3. (Cited on pages 113, 168, and 169.)

Berdondini, L., Imfeld, K., Maccione, A., Tedesco, M., Neukom, S., Koudelka-Hep, M., and Martinoia, S. (2009a). Active pixel sensor array for high spatio-temporal resolution electrophysiological recordings from single cell to large scale neuronal networks. *Lab Chip*, 9:2644–51. (Cited on pages 11, 17, and 45.)

Berdondini, L., Massobrio, P., Chiappalone, M., Tedesco, M., Imfeld, K., Maccione, A., Gandolfo, M., Koudelka-Hep, M., and Martinoia, S. (2009b). Extracellular recordings from locally dense microelectrode arrays coupled to dissociated cortical cultures. *J Neurosci Meth*, 177:386–396. (Cited on pages 10, 11, 17, 25, and 45.)

Berdondini, L., van der Wal, P., Guernat, O., de Rooij, N., Koudelka-Hep, M., Seitz, P., Kaufman, R., Metzler, P., Blanc, N., and Rohr, S. (2005). High-density electrode array for imaging in vitro electrophysiology activity. *Biosens Bioelectron*, 21(1):167–

174. (Cited on pages 11, 15, 17, 45, and 82.)

Berkes, P., Orban, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment science. *Science*, 331:83–87. (Cited on page 160.)

Bettencourt, L., Stephens, G., Ham, M., and Gross, G. (2007). Functional structure of cortical neuronal networks grown in vitro. *Phys Rev E*, 75:021915. (Cited on pages 9, 44, 85, 160, 161, and 172.)

Bi, G.-q. and Poo, M.-m. (1999). Distributed synaptic modification in neural networks induced by patterned stimulation. *Nature*, 401:792–796. (Cited on pages 1, 2, 9, and 44.)

Blankenship, A. and Feller, M. (2010). Mechanisms underlying spontaneous patterned activity in developing neural circuits. *Nat Rev Neurosci*, 11:18–29. (Cited on page 5.)

Brette, R. (2015). Philosophy of the spike: rate-based vs. spike-based theories of the brain. *Frontiers in systems neuroscience*, 9:151. (Cited on pages 155 and 171.)

Brown, K. and Sethna, J. (2003). Statistical mechanical approaches to models with many poorly known parameters. *Phys Rev E*, 68:021904. (Cited on pages 89, 90, 91, 108, 135, 136, 137, and 172.)

Brush, S. G. (1967). History of the lenz-ising model. *Rev. Mod. Phys.*, 39:883–893. (Cited on page 47.)

Buzsáki, G. and Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nat Rev Neurosci*, 15:264–278. (Cited on pages 44, 76, 122, and 168.)

Carmena, J., Lebedev, M., Henriquez, C., and Nicolelis, M. (2005). Stable ensemble performance with single-neuron variability during reaching movements in primates. *J Neurosci*, 25:10712–6. (Cited on pages 2, 3, 140, and 167.)

Chestek, C., Batista, A., Santhanam, G., Yu, B., Afshar, A., Cunningham, J., Gilja, V., Ryu, S., Churchland, M., and Shenoy, K. (2007). Single-neuron stability during repeated reaching in macaque premotor cortex. *J Neurosci*, 27:10742–50. (Cited on pages 2, 3, and 140.)

Chiappalone, M., Bove, M., Vato, A., Tedesco, M., and Martinoia, S. (2006). Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development. *Brain Res*, 1093:41–53. (Cited on pages 10, 11, 15, 44, and 45.)

Chklovskii, D. B., Mel, B., and Svoboda, K. (2004). Cortical rewiring and information storage. *Nature*, 431(7010):782–788. (Cited on page 1.)

Chu, C., Chien, P., and Hung, C. (2014a). Multi-electrode recordings of ongoing activity and responses to parametric stimuli in macaque v1. *CRCNS.org*. (Cited on pages 141 and 154.)

Chu, C., Chien, P., and Hung, C. (2014b). Tuning dissimilarity explains short distance decline of spontaneous spike correlation in macaque v1. *Vis Res*, 96:113–132. (Cited on page 141.)

Colombi, I., Tinarelli, F., Pasquale, V., Tucci, V., and Chiappalone, M. (2016). A simplified in vitro experimental model encompasses the essential features of sleep. *Frontiers in Neuroscience*, 10. (Cited on pages 139, 155, 160, and 169.)

Corner, M. (2008). Spontaneous neuronal burst discharges as dependent and independent variables in the maturation of cerebral cortex tissue cultured *in vitro*: A review of activity-dependent studies in live 'model' systems for the development of intrinsically generated bioelectric slow-wave sleep patterns. *Brain Res Rev*, 59:221–244. (Cited on pages 9, 11, 30, 45, 139, 154, and 160.)

Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons. (Cited on page 54.)

Crooks, G. E. (2007). Measuring thermodynamic length. *Phys Rev Lett*, 99:100602–6. (Cited on pages 47, 81, 92, 93, 94, 96, and 98.)

Daniels, B., Chen, Y.-J., Sethna, J., Gutenkunst, R., and Myers, C. (2008). Sloppiness, robustness, and evolvability in systems biology. *Curr Opin Biotech*, 19:389–395. (Cited on pages 5, 89, 90, 91, 92, 105, 108, 135, 136, and 137.)

Davis, G. (2006). Homeostatic control of neural activity: From phenomenology to molecular design. *Annu Rev Neurosci*, 29:307–323. (Cited on pages 3 and 5.)

Erickson, J., Tooker, A., Tai, Y.-C., and Pine, J. (2008). Caged neuron mea: A system for long-term investigation of cultured neural network connectivity. *J Neurosci Meth*, 175:1–16. (Cited on page 28.)

Fitzsimonds, R., Song, H., and Poo, M. (1997). Propagation of activity-dependent synaptic depression in simple neural networks. *Nature*, 388:439–448. (Cited on pages 1, 2, 9, and 28.)

Fusi, S. and Abbott, L. (2007). Limits on the memory storage capacity of bounded synapses. *Nat Neurosci*, 10:485–93. (Cited on page 2.)

Ganguly, K. and Carmena, J. M. (2009). Emergence of a stable cortical map for neuroprosthetic control. *PLoS biology*, 7(7):1508. (Cited on pages 1, 3, 140, and 167.)

Golowasch, J., Casey, M., Abbot, L., and Marder, E. (1999). Neuronal stability from activity-dependent regulation of neuronal conductances. *Neural Comput*, 11:1079–1096. (Cited on pages 4, 166, and 172.)

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., and Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3(10):e189. (Cited on page 91.)

Han, E. and Stevens, C. (2009). Development regulates a switch between post- and presynaptic strengthening in response to activity deprivation. *PNAS*, 106(26):12159–12164. (Cited on pages 9 and 10.)

Hinard, V., Mikhail, C., Pradervand, S., Curie, T., Houtkooper, R. H., Auwerx, J., Franken, P., and Tafti, M. (2012). Key electrophysiological, molecular, and metabolic signatures of sleep and wakefulness revealed in primary cortical cultures. *The Journal of neuroscience*, 32(36):12506–12517. (Cited on pages 9, 139, 140, 155, 160, and 169.)

Hobson, J. A., Pace-Schott, E. F., and Stickgold, R. (2000). Dreaming and the brain: toward a cognitive neuroscience of conscious states. *Behavioral and Brain Sciences*, 23(06):793–842. (Cited on page 170.)

Holtmaat, A. J. G., Trachtenberg, J., Wilbrecht, L., Shepherd, G., Zhang, X., Knott, G., and Svoboda, K. (2005). Transient and persistent dendritic spines in the neocortex in vivo. *Neuron*, 45:279–91. (Cited on pages 1 and 2.)

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *PNAS*, 79:2554–2558. (Cited on pages 47, 48, 55, 75, 86, and 161.)

Huber, D., Gutnisky, D., Peron, S., O'Connor, D., Wiegert, J., Tian, L., Oertner, T., Looger, L., and Svoboda, K. (2012). Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature*, 484:473–478. (Cited on pages 2 and 3.)

Hung, C., Cui, D., Chen, Y.-p., Lin, C.-p., and Levine, M. (2014). Correlated activity supports efficient cortical processing. *Front Comput Neurosci*, 8:1–16. (Cited on pages 167 and 168.)

Hurley, N. and Rickard, S. (2009). Comparing measures of sparsity. *IEEE T Inform Theory*, 55:4723–4741. (Cited on page 102.)

Izhikevich, E. (2003). Simple model of spiking neurons. *IEEE T Neu Net*, 14(6):1569–1572. (Cited on page 172.)

Kamioka, H., Maeda, E., Jimbo, Y., Robinson, H. P., and Kawana, A. (1996). Spontaneous periodic synchronized bursting during formation of mature patterns of connections in cortical cultures. *Neuroscience letters*, 206(2):109–112. (Cited on pages 2, 9, 10, 11, 15, 30, 44, and 45.)

Lever, C., Wills, T., Cacucci, F., Burgess, N., and O'Keefe, J. (2002). Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature*, 416(6876):90–94. (Cited on pages 2 and 166.)

Lütcke, H., Margolis, D. J., and Helmchen, F. (2013). Steady or changing? long-term monitoring of neuronal population activity. *Trends in neurosciences*, 36(7):375–384. (Cited on pages 14, 44, 78, 140, and 170.)

Machta, B., Chachra, R., Transtrum, M., and Sethna, J. (2013). Parameter space compression underlies emergent theories and predictive models. *Science*, 342:604–7. (Cited on pages 89, 91, 94, 96, 100, 101, 104, 105, 108, 112, 135, 136, 137, and 168.)

Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S., and Bethge, M. (2009). Generating spike trains with specified correlation coefficients. *Neural Computation*, 21(2):397–423. (Cited on pages 51 and 70.)

Macke, J. H., Murray, I., and Latham, P. E. (2013). Estimation bias in maximum entropy models. *Entropy*, 15(8):3109–3129. (Cited on pages 61 and 176.)

Maletic-Savatic, M., Malinow, R., and Svoboda, K. (1999). Rapid dendritic morphogenesis in ca1 hippocampal dendrites induced by synaptic activity. *Science*, 283:1923–7. (Cited on page 1.)

Marder, E. and Goaillard, J.-M. (2006). Variability, compensation and homeostasis in neuron and network function. *Nat Rev Neurosci*, 7:563–575. (Cited on pages 3, 5, 159, and 166.)

Marder, E. and Prinz, A. (2002). Modeling stability in neuron and network function: the role of activity in homeostasis. *Bioessays*, 24:1145–1154. (Cited on page 3.)

Margolis, D., Lütcke, H., Schulz, K., Haiss, F., Weber, B., Kügler, S., Hasan, M., and Helmchen, F. (2012). Reorganization of cortical population activity imaged throughout long-term sensory deprivation. *Nat Neurosci*, 15:1539–46. (Cited on pages 1, 2, 3, 44, 45, and 167.)

Marom, S. and Shahaf, G. (2002). Development, learning and memory in large random networks of cortical neurons: lessons beyond anatomy. *Quarterly reviews of biophysics*, 35(01):63–87. (Cited on pages 2, 9, 10, 11, 44, 45, 76, and 139.)

Minerbi, A., Kahana, R., Goldfeld, L., Kaufman, M., Marom, S., and Ziv, N. (2009). Long-term relationships between synaptic tenacity, synaptic remodelling, and network activity. *PLoS Biol*, 7(6):e1000136(1–20). (Cited on pages 1, 2, 9, 11, 13, 17, 38, 46, and 160.)

Nakayama, K., Kiyosue, K., and Taguchi, T. (2005). Diminished neuronal activity increases neuron-neuron connectivity underlying silent synapse formation and the rapid conversion of silent to functional synapses. *The Journal of neuroscience*, 25(16):4040–4051. (Cited on pages 1, 10, and 45.)

Ohiorhenuan, I., Mechler, F., Purpura, K., Schmid, A., Hu, Q., and Victor, J. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature*, 466:617–622. (Cited on pages 47, 50, 51, 52, 84, and 85.)

Okabe, S., Kim, H.-D., Miwa, A., Kuriu, T., and Okado, H. (1999). Continual remodeling of postsynaptic density and its regulation by synaptic activity. *Nature Neuroscience*, 2(9):804–811. (Cited on pages 1, 9, 10, 11, 17, and 38.)

Panas, D., Amin, H., Maccione, A., Muthmann, O., van Rossum, M., Berdondini, L., and Hennig, M. H. (2015). Sloppiness in spontaneously active neuronal networks. *The Journal of Neuroscience*, 35(22):8480–8492. (Cited on page 143.)

Parrondo, J. M., Horowitz, J. M., and Sagawa, T. (2015). Thermodynamics of information. *Nature Physics*, 11(2):131–139. (Cited on page 47.)

Prinz, A., Bucher, D., and Marder, E. (2004). Similar network activity from disparate ciruit parameters. *Nat Neurosci*, 7(12):1345–1352. (Cited on page 166.)

Quiroga, R., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput*, 16(8):1661–1687. (Cited on page 142.)

Rechtschaffen, A. and Kales, A. (1968). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. (Cited on page 170.)

Roudi, Y., Nirenberg, S., and Latham, P. (2009a). Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS Comput Biol*, 5:e1000380. (Cited on pages 49 and 58.)

Roudi, Y., Tyrcha, J., and Hertz, J. (2009b). Ising model for neural data: model quality and approximate methods for extracting functional connectivity. *Physical Review E*, 79(5):051915. (Cited on pages 58, 85, and 164.)

Schneidman, E., Berry II, M., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007–1012. (Cited on pages 47, 48, 49, 50, 51, 52, 58, 60, 61, 63, 66, 67, 69, 70, 71, 73, 77, 79, 82, 84, 86, 99, 125, 126, 160, 161, and 162.)

Schneidman, E., Still, S., Berry II, M., and Bialek, W. (2003). Network information and connected correlations. *Phys Rev Lett*, 91(23):238701(1–4). (Cited on page 60.)

Schroeter, M. S., Charlesworth, P., Kitzbichler, M. G., Paulsen, O., and Bullmore, E. T. (2015). Emergence of rich-club topology and coordinated dynamics in development of hippocampal functional networks in vitro. *The Journal of Neuroscience*, 35(14):5459–5470. (Cited on pages 76, 77, 160, 161, and 172.)

Schwarz, D. A., Lebedev, M. A., Hanson, T. L., Dimitrov, D. F., Lehew, G., Meloy, J., Rajangam, S., Subramanian, V., Ifft, P. J., Li, Z., et al. (2014). Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. *Nature methods*, 11(6):670–676. (Cited on page 170.)

Segev, R., Baruchi, I., Hulata, E., and Ben-Jacob, E. (2004). Hidden neuronal correlations in cultured networks. *Phys Rev Lett*, 92:118102. (Cited on pages 10, 11, 44, and 76.)

Segev, R., Shapira, Y., Benveniste, M., and Ben-Jacob, E. (2001). Observations and modeling of synchronized bursting in two-dimensional neural networks. *Physical Review E*, 64(1):011920. (Cited on pages 10, 30, and 44.)

Shahaf, G., Eytan, D., Gal, A., Kermany, E., Lyakhov, V., Zrenner, C., and Marom, S. (2008). Order-based representation in random networks of cortical neurons. *PLoS Comput. Biol*, 4(11):e1000228. (Cited on pages 9, 10, 12, 44, and 171.)

Shew, W., Yang, H., Yu, S., Roy, R., and Plenz, D. (2011). Information capacity and transmission are maximized in balanced cortical networks with cortical avalanches. *J Neurosci*, 31(1):15595–15600. (Cited on pages 113 and 169.)

Shlens, J., Field, G., Gauthier, J., Grivich, M., Petrusca, D., Sher, A., Litke, A., and Chichilnisky, E. (2006). The structure of multi-neuron firing patterns in primate retina. *J Neurosci*, 26(32):8254–8266. (Cited on pages 47, 49, 52, 58, 60, 61, 63, 71, 82, 86, 160, and 162.)

Shoham, S., O'Connor, D., and Segev, R. (2006). How silent is the brain: is there a "dark matter" problem in neuroscience? *J Comp Physiol A*, 192(8):777–784. (Cited on pages 45, 140, and 170.)

Slomowitz, E., Styr, B., Vertkin, I., Milshtein-Parush, H., Nelken, I., Slutsky, M., and Slutsky, I. (2015). Interplay between population firing stability and single neuron dynamics in hippocampal networks. *eLIFE*, 4(e04378):1–21. (Cited on pages 1, 2, 11, 13, 25, 38, 44, 46, 160, and 171.)

Steriade, M., Nuñez, A., and Amzica, F. (1993a). Intracellular analysis of relations between the slow (< 1 hz) neocortical oscillation and other sleep rhythms of the electroencephalogram. *The Journal of Neuroscience*, 13(8):3266–3283. (Cited on page 140.)

Steriade, M., Nunez, A., and Amzica, F. (1993b). A novel slow (< 1 hz) oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *The Journal of Neuroscience*, 13(8):3252–3265. (Cited on page 139.)

Sweeney, Y., Kotaleski, J. H., and Hennig, M. H. (2015). A diffusive homeostatic signal maintains neural heterogeneity and responsiveness in cortical networks. *PLoS Comput Biol*, 11(7):e1004389. (Cited on page 4.)

Tang, A., Jackson, D., Hobbs, J., Chen, W., Smith, J., Patel, H., Prieto, A., Petrusca, D., Grivich, M., Sher, A., Hottowy, P., Dabrowski, W., Litke, A., and Beggs, J. (2008). A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J Neurosci*, 28(2):505–518. (Cited on pages 47, 50, 51, 52, 56, 57, 58, 60, 61, 63, 67, 69, 70, 71, 73, 77, 78, 82, 84, 86, and 162.)

Tkacik, G., Schneidman, E., Berry, I., Michael, J., and Bialek, W. (2006). Ising models for networks of real neurons. *arXiv preprint q-bio/0611072*.

Trachtenberg, J., Chen, B., Knott, G., Feng, G., Sanes, J., Welker, E., and Svoboda, K. (2002). Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, 420:788–94. (Cited on pages 1, 2, 140, and 166.)

Transtrum, M. K., Machta, B. B., Brown, K. S., Daniels, B. C., Myers, C. R., and Sethna, J. P. (2015). Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of chemical physics*, 143(1):010901. (Cited on pages 5, 89, 90, 91, 92, 105, 107, 108, 113, 135, 136, 137, 159, 172, and 173.)

Turrigiano, G. (1999). Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends Neurosci*, 22(5):221–227. (Cited on page 3.)

Turrigiano, G. (2011). Too many cooks? intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement. *Annu Rev Neurosci*. (Cited on pages 3, 4, and 159.)

Turrigiano, G. and Nelson, S. (2000). Hebb and homeostasis in neuronal plasticity. *Curr Opp Neurobiol*, 10(3):358–364. (Cited on pages 2, 3, and 4.)

Verderio, C., Coco, S., Pravettoni, E., Bacci, A., and Matteoli, M. (1999). Synaptogenesis in hippocampal cultures. *Cell Mol Life Sci*, 55:1448–62. (Cited on pages 9 and 10.)

Wagenaar, D., Pine, J., and Potter, S. (2006). An extremely rich repertoire of bursting patterns during the development of cortical cultures. *BMC Neurosci*, 7(11). (Cited on pages 10, 11, 12, 15, 22, 30, 31, 44, and 45.)

Yu, S., Huang, D., Singer, W., and Nikolić, D. (2008). A small world of neuronal synchrony. *Cerebral cortex*, 18(12):2891–2901. (Cited on pages 47, 49, 52, 58, 60, 61, 63, 67, 71, 73, 75, 76, 82, 86, 156, 160, 161, 162, and 172.)

Yu, S., Yang, H., Nakahara, H., Santos, G., Nikolic, D., and Plenz, D. (2011). Higher-order interactions characterized in cortical activity. *J Neurosci*, 31(48):17514–17526. (Cited on pages 50, 51, 52, 58, 60, 61, 67, 70, 84, 85, 86, and 172.)

Zeng, H.-L., Alava, M., Aurell, E., Hertz, J., and Roudi, Y. (2013). Maximum likelihood reconstruction for ising models with asynchronous updates. *Physical review letters*, 110(21):210601. (Cited on page 171.)

Ziv, Y., Burns, L., Cocker, E., Hamel, E., Ghosh, K., Kitch, L., Gamal, A., and Schnitzer, M. (2013). Long-term dynamics of ca1 hippocampal place codes. *Nat Neurosci*, 16:264–266. (Cited on pages 1, 2, 3, 140, and 167.)

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography *"The Elements of Typographic Style"*. `classicthesis` is available for both LaTeX and LyX:

http://code.google.com/p/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of May 18, 2017 (`classicthesis` version 4.0).